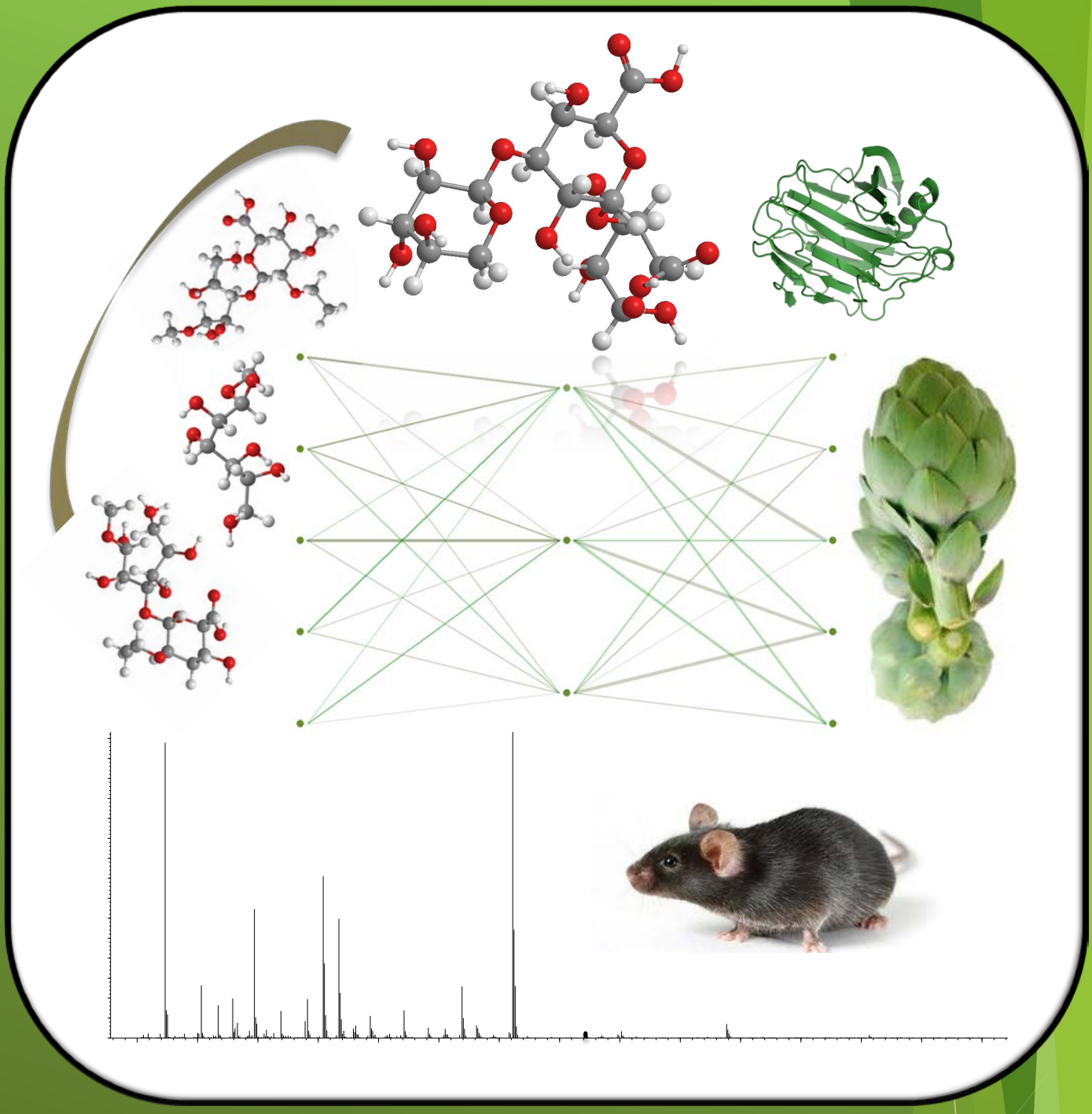


ENZYMATIC OBTAINMENT OF PECTIN AND PECTIC OLIGOSACCHARIDES FROM ARTICHOKE BY-PRODUCTS

STRUCTURAL CHARACTERISATION AND FUNCTIONAL EVALUATION THROUGH MACHINE LEARNING

CARLOS SABATER SÁNCHEZ

PhD THESIS 2019

UAM UNIVERSIDAD AUTONOMA DE MADRID

CSIC CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

CIAL INSTITUTO DE INVESTIGACIÓN EN CIENCIAS DE LA ALIMENTACIÓN

UNIVERSIDAD AUTÓNOMA DE MADRID

Facultad de Ciencias

Departamento de Química-Física Aplicada

Enzymatic obtainment of pectin and pectic oligosaccharides from artichoke by-products. Structural characterization and functional evaluation through machine learning

CARLOS SABATER SÁNCHEZ

PhD Thesis, International Mention

Madrid, 2019

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

INSTITUTO DE INVESTIGACIÓN EN CIENCIAS DE LA ALIMENTACIÓN

UNIVERSIDAD AUTÓNOMA DE MADRID

Facultad de Ciencias

Departamento de Química-Física Aplicada

Enzymatic obtainment of pectin and pectic oligosaccharides from artichoke by-products. Structural characterization and functional evaluation through machine learning

CARLOS SABATER SÁNCHEZ

Thesis submitted in fulfilment of the requirements for the degree of doctor at Autonomous University of Madrid (UAM)
Madrid, 2019

Thesis supervisors:
Dr. Nieves Corzo Sánchez
Dr. Antonia Montilla Corredera
(Spanish National Research Council, CIAL-CSIC)

Nieves Corzo Sánchez and **Antonia Montilla Corredera**, Research Professor and Tenured Scientist of the Spanish National Research Council (CSIC) at the Institute of Food Science Research (CIAL),

CERTIFY:

That Carlos Sabater Sánchez has performed, under their supervision, the research work entitled: "**Enzymatic obtainment of pectin and pectic oligosaccharides from artichoke by-products. Structural characterization and functional evaluation through machine learning**". This work memory is submitted in fulfillment of the requirements for the degree of doctor at the Autonomous University of Madrid (UAM).

Madrid, 4 October 2019

Fdo.: Dra. Nieves Corzo Sánchez

Fdo.: Dra. Antonia Montilla Corredera

Agradecimientos

En primer lugar, me gustaría dar las gracias a mis directoras de Tesis, Nieves y Toñi, por toda la confianza, el apoyo y el cariño que me han mostrado a lo largo de estos años, siendo un referente para mí. Ha sido un auténtico placer trabajar a su lado, y sus palabras y consejos me han dado ánimos para seguir adelante y han aumentado mi interés por esta profesión. En este sentido, también quiero dar las gracias al resto de miembros del grupo de Carbohidratos, Mar, Javi, Agustín, Oswaldo, a mis compañeros de laboratorio (Nerea, Álvaro, Paloma, Ana, Cristina, Agustina, Mayte, Pablo), así como a otros compañeros que han trabajado con nosotros (Ana, Celia, Paula, Estrella, Berta, Andrés, Raquel, Javi, Félix, Inés, Pablo). También quiero aprovechar para dar las gracias a los compañeros de otros grupos de investigación, a algunos de los cuales conozco desde que éramos estudiantes de Grado y Máster, y a todo el personal del CIAL, siempre dispuesto a ayudar haciendo mi trabajo más fácil.

Agradezco al Ministerio de Educación y Formación Profesional la beca concedida (FPU14/03619), que me ha permitido realizar mis estudios de doctorado. También quiero dar las gracias a la Universidad Autónoma de Madrid (UAM), y en especial a Marín, a Tiziana y al resto de profesores del área de alimentos, a los que conozco desde que ingresé en la universidad, y gracias a los cuales he descubierto mi vocación investigadora.

La realización de estancias en otros centros me ha permitido conocer grupos de investigación diferentes, ayudándome a continuar con mi formación y de los cuales guardo muy buenos recuerdos. Es por ello que quiero dar las gracias:

A Gabi y a todos los miembros de su grupo en el CERELA, por haberme acogido con tanta amabilidad y por haberme regalado una gran experiencia.

Al profesor Mackie y al profesor Harding, por haberme recibido en sus grupos de investigación, y en especial a Amelia por apoyarme tanto a lo largo de mis estancias en Leeds. También quiero dar las gracias al resto de compañeros de estos grupos de investigación, Vlad, Can Can, Fadel, así como a los compañeros de los otros grupos de la School of Food Science con los que he compartido muy buenos momentos.

A la profesora Utrilla y al profesor Gálvez, así como a todos los miembros de su grupo, José Alberto, Teresa, Antonio, Laura, Patri y María Jesús, por haberme ayudado tanto con mis ratones, y por haber hecho que mi estancia sea muy agradable.

Al profesor Margolles por todo el interés que ha mostrado en nuestra investigación, por toda su amabilidad y su ayuda, y por motivarme a continuar con mi carrera investigadora.

Por otra parte, quiero dar las gracias a mi familia. A mi hermano Víctor y a mis padres, por ser un ejemplo a seguir en todos los aspectos y por haberme enseñado el valor del esfuerzo y la perseverancia desde pequeño. Si he llegado hasta aquí ha sido gracias a vosotros. También quiero agradecer a Mario y Álvaro la larga amistad que nos une desde el colegio y todo lo que hemos vivido juntos. Por último, quiero terminar dando las gracias a Lucía por toda la confianza y la complicidad, por animarme a dar siempre lo mejor de mí mismo y por ser el amor de mi vida.

Gracias.

CONTENTS

Abbreviation list

Abbreviation: Definition
AloLa: Allolactose
ANN: Artificial neural network
AP: Artichoke pectin
APOS: Artichoke pectic oligosaccharides
APwA: Modified artichoke pectin without arabinose
APwG: Modified artichoke pectin without galactose
Ara: Arabinose
BLR: Boosted logistic regression
CAN: Cellulase from *Aspergillus niger*
CFM-ID: Competitive fragmentation modelling for metabolite identification
CIT: Sodium citrate
CNN: Convolutional neural network
CP: Citrus pectin
CPOS: Citrus pectic oligosaccharides
DM: Degree of methyl-esterification
DNS: Dinitrosalicylic acid
DP: Degree of polymerization
DSS: Dextran sulfate sodium
DWT: Discrete wavelet transform
E: Enzymatic treatment
ELSD: Evaporative light scattering detector
Endo-PG: Endo-polygalacturonase
Exo-PG: Exo-polygalacturonase
FOS: Fructo-oligosaccharides
Fru: Fructose
FT-IR: Fourier transform infrared spectroscopy
GaBio: Galactobiose
Gal: Galactose
GalA(Acet): Acetyl-esterified galacturonic acid
GalA(Met): Methyl-esterified galacturonic acid
GalA: Galacturonic acid
GC-EI-MS: Gas chromatography – electron impact – mass spectrometry
GC-FID: Gas chromatography – flame ionization detector
Glc: Glucose
GLM-NET: Generalized linear model – elastic net
GOS: Galacto-oligosaccharides
HG: Homogalacturonan
HILIC: Hydrophilic interaction chromatography
HMDS: Hexamethyldisilazane
HMP: High methyl-esterified pectin
HPAEC-PAD: High performance anion exchange chromatography – pulsed amperometric detector
HPSEC: High performance size exclusion chromatography

IBD: Inflammatory bowel disease

IS: Internal standard

La: Lactose

LMP: Low methyl-esterified pectin

MALDI-TOF-MS: Matrix assisted laser desorption ionization – time of flight – mass spectrometry

Man: Mannose

MLP: Multilayer perceptron

MS: Mass spectrometry

M_w: molecular weight

NB: Naïve Bayes classifier

NDOs: Non-digestible oligosaccharides

NIT: Nitric acid

PCA: Principal components analysis

PL: Pectin lyase

PLS-DA: Partial least squares discriminant analysis

PME: Pectin methyl esterase

POS: Pectic oligosaccharides

PUO: Pectinex ultra olio

qPCR: Quantitative polymerase chain reaction

RF: Random forest

RG-I: Rhamnogalacturonan-I

RG-II: Rhamnogalacturonan-II

Rha: Rhamnose

RID: Refractive index detector

SCFAs: Short chain fatty acids

TFA: Trifluoroacetic acid

TMSO: Trimethyl sililated oximes

US: Ultrasounds

US+E: Ultrasounds in combination with enzymes

Xyl: Xylose

Summary

1. SUMMARY

There is an overall interest in finding alternative ways of revalorization of vegetable processing by-products, which could be used as a source of biologically active ingredients. Most pectins used in the industry are extracted from citrus peels, apple and sugar beet pulp. However, efforts have been made on studying alternative sources of pectin that may exhibit different technological and biological properties. It has been demonstrated that artichoke (*Cynara scolymus*) by-products are a good source of functional ingredients. This PhD Thesis aims at extracting and characterizing artichoke pectin, as well as evaluating its enzymatic depolymerization to obtain pectic oligosaccharides (POS) mixtures. Then, the use of these products as functional ingredients was assessed by studying their behavior when incorporated in a dairy matrix. In addition, the anti-inflammatory activity *in vivo* of some of these products has been evaluated in a DSS-induced model of mice colitis. For this purpose, artichoke pectin extraction using a cellulase preparation, Celluclast, has been optimized leading to high yields, and could be used as a food additive (E-440) because presented a high content of galacturonic acid (> 65%). Other structural characteristics of artichoke pectin were high neutral sugars content, especially arabinose, multimodal distribution of molecular weight, and low degree of methyl-esterification. Enzymatic extraction of pectin has been compared to traditional extraction using sodium citrate and nitric acid, as well as ultrasound-assisted extraction in combination with cellulase, which could be used to obtain artichoke pectin at shorter times achieving good yields and without important structural modification. On the other hand, POS obtainment from artichoke pectin using enzymes with different activities has been studied, and compared to POS from highly methyl-esterified citrus and apple pectins. Cellulase from *Aspergillus niger* and Pectinex Ultra Olio formed large amounts of artichoke POS. To elucidate the mass spectra of pectin and POS chains when no reference standard or spectra library is available, a computational data analysis strategy based on the combination of machine learning algorithms and *in silico* fragmentation has been developed. This computational method has been also used to characterize POS structures released in small amounts during the manufacture of functional yogurts containing citrus and artichoke pectin and POS. Finally, *in vivo* studies found that artichoke pectin reduced pro-inflammatory

cytokine expression showing an enhanced bioactivity compared to high methyl esterified citrus pectin. Expression profiles were established through artificial neural networks and it was observed that galactose content was especially relevant to preserve anti-inflammatory activity of pectin.

Resumen

2. RESUMEN

Hay un interés general por encontrar nuevas formas de revalorización de los subproductos generados durante el procesamiento de vegetales, pudiendo ser usados como fuente de ingredientes biológicamente activos. La mayoría de las pectinas utilizadas en la industria se extraen a partir de cítricos, manzana y remolacha azucarera. Sin embargo, se están investigando otras fuentes alternativas de pectina que puedan presentar diferentes propiedades tecnológicas y biológicas. Se ha demostrado que los subproductos de alcachofa (*Cynara scolymus*) son una buena fuente de ingredientes funcionales. Esta Tesis Doctoral tiene por objetivo la extracción y la caracterización de la pectina de alcachofa, así como el estudio de la despolimerización enzimática de la misma para obtener mezclas de oligosacáridos pécticos (POS). Además, se ha evaluado el posible uso de estos productos como ingredientes funcionales estudiando tanto su comportamiento en una matriz láctea como su actividad anti-inflamatoria *in vivo* en un modelo de colitis inducida por DSS en ratones. Con este fin, se ha optimizado la extracción de pectina de alcachofa utilizando un preparado de celulasa, Celluclast, obteniéndose altos rendimientos. Esta pectina podría ser utilizada como aditivo alimentario (E-440) por su elevado contenido en ácido galacturónico (> 65%). Algunas de sus características estructurales fueron su alto contenido en azúcares neutros, especialmente arabinosa, su distribución multimodal de masas moleculares y su bajo grado de metil-esterificación. Se ha comparado la extracción enzimática con la extracción convencional mediante citrato sódico y ácido nítrico, y con la extracción asistida por ultrasonidos en combinación de una celulasa. Ésta última podría ser utilizada para obtener pectina de alcachofa a tiempos más cortos lográndose buenos rendimientos y sin una modificación estructural importante. Por otra parte, se ha estudiado la obtención de POS a partir de pectina de alcachofa utilizando enzimas con diferentes actividades, y se han comparado los perfiles de POS con aquellos obtenidos a partir de pectinas de cítricos y manzana de alto grado de metil-esterificación. Las enzimas Celulasa de *Aspergillus niger* y Pectinex Ultra Olio dieron lugar a elevados rendimientos de POS. Para elucidar los perfiles estructurales de las cadenas de pectinas y POS a partir de sus espectros de masas, se ha desarrollado una nueva estrategia de análisis de datos basada en la combinación de algoritmos de aprendizaje automático y la

fragmentación *in silico*. Este método computacional también se ha utilizado para caracterizar diversas estructuras de POS liberadas en pequeñas cantidades durante la elaboración de yogures funcionales con pectina y POS de cítricos y alcachofa. Por último, los estudios *in vivo* revelaron que la pectina de alcachofa reduce la expresión de citoquinas pro-inflamatorias, mostrando una bioactividad mejorada con respecto a la pectina de cítricos. Se han establecido perfiles de expresión mediante redes neuronales artificiales y se observó que el contenido en galactosa de la pectina resultó ser especialmente relevante para conservar su actividad anti-inflamatoria.

General introduction

3. GENERAL INTRODUCTION

3.1. Functional foods

Nowadays, there is a consciousness-raising about the management of industrial by-products due to their high environmental impact; because of this, the food industry is trying to develop new ways of reuse and revalorization. Traditional uses of agro-industrial by-products include animal feeding and compost production. However, these substrates could be used as natural sources to isolate biologically-active compounds for drug and functional food development (Aruwa et al., 2018). To determine the potential applications of industrial wastes, each by-product needs to be characterized.

Several definitions have been proposed for functional foods although there is no general consensus. The European Commission's Concerted Action on Functional Food Science in Europe (FuFoSE), coordinated by International Life Science Institute (ILSI) of Europe proposed the following definition: "a food product can only be considered functional if together with the basic nutritional impact it has beneficial effects on one or more functions of the human organism thus either improving the general and physical conditions or/and decreasing the risk of the evolution of diseases. The amount of intake and form of the functional food should be as it is normally expected for dietary purposes. Therefore, it could not be in the form of pill or capsule just as normal food form" (Diplock et al., 1999). This is one of the first and most widespread definitions. One recent study defined functional foods as "natural or processed foods that contain biologically active compounds; which, in defined, effective, and non-toxic amounts, provide a clinically proven and documented health benefit utilizing specific biomarkers for the prevention, management, or treatment of chronic disease or its symptom" (Gur et al., 2018). This may include unmodified natural foods that have demonstrated to provide health benefits, as well as conventional foods to which chemical composition has been modified to subserve physiological roles. According to it, different strategies may be followed for the development of functional foods including the addition of a food natural ingredient, the removal of a food compound, or the increase on the concentration or bioavailability of a functional component. These positive effects may

improve consumer's physical well-being and/or reducing the risk of several chronic diseases such as cardiovascular and intestinal diseases or obesity. In addition, functional food must maintain the physical structure equivalent to its non-functional analogue (European Commision, 2010). High-value functional food ingredients used by the industry include phytochemical, probiotics, prebiotic carbohydrates, polyunsaturated fatty acids and bioactive peptides/proteins (Fernandes et al., 2019).

3.2. Bioactive carbohydrates

Carbohydrates can be considered as digestible and non-digestible (poly- and oligosaccharides NDOs) depending on their resistance to the hydrolytic activity of the human digestive enzymes. The polysaccharides are typically long polymeric carbohydrate chains containing up to several hundred thousand monomeric units; while, most oligosaccharides contain 2 to 10 sugar moieties, although the degree of polymerization (DP) could go up to 60 for some, like chicory inulin or down to 2 for some NDOs like lactulose (Sweenen et al. 2006).

NDOs show interesting properties and they were proposed as "prebiotics" for the first time by Delzenne and Roberfroid (1994) which were described as food ingredients that can modify the composition of endogenous gut microflora. Some of these NDOs can be constituents of common foods. Later, Gibson and Roberfroid (1995) defined for the first time a prebiotic as "non-digestible food ingredient that beneficially affects the host by selectively stimulating the growth and /or activity of one or a limited number of bacteria in the colon, and thus improves host health". This definition has changed over time being the most recent definition the one proposed by the International Scientific Association for Probiotics and Prebiotics consensus (ISAPP) in 2017: "a substrate that is selectively utilized by host microorganisms conferring a health benefit". This definition expands the concept of prebiotics to include non-carbohydrate substances such as phenolic compounds or polyunsaturated fatty acids, and their applications to body sites other than the gastrointestinal tract (such as skin or vagina) and diverse categories other than food. Compared to the direct administration of probiotic microorganisms, the prebiotic target are bacteria present already in the host and they stimulate its growth.

Some health benefits associated with the dietary fibre consumption include an increase in faecal dry weight excretion due to bulk effect and water absorption, in the case of non-fermentable fibre, or by the increased number of bacteria resulting in constipation relief, inhibition of diarrhoea related to the possible inhibitory effect of bifidobacteria (stimulated by NDOs) on the adhesion of harmful bacteria to the epithelial surfaces and a reduction of cancer risk related to an increase in cellular immunity produced by cell wall and extra-cellular components of bifidobacteria (Enam & Mansell, 2019; Sanders et al., 2019).

Extensive research has been carried out to identify potentially prebiotic compounds. A growing number of *in vitro* and *in vivo* studies show that prebiotics could induce beneficial physiological effects in the colon because contribute towards reduction of the risk of dysbiosis and intestinal and systemic pathologies (Roberfroid, et al., 2010). Other beneficial effects of prebiotic consumption involve an increase calcium bioavailability and improvement of mineral retention, modulating the rate of bone loss associated with ageing, as well as an immunostimulatory activity in allergic response or during chronic inflammatory diseases, reducing pro-inflammatory cytokine secretion and ameliorating the symptoms of inflammatory bowel disease (IBD). Similarly, positive effects of prebiotics have been reported for several paediatric disorders and exert preventive effects on gastro-intestinal infections. In addition, prebiotic consumption leads to a reduction cholesterol absorption and may reduce the risk of colon cancer by modulating the colonic microbiota involved in the etiology of cancer (Brosseau et al., 2019; Khangwal & Shukla, 2019; Roberfroid et al., 2010). It should be considered that bacterial metabolism of unabsorbed dietary residues and endogenous secretions are origin of several genotoxic and tumor-promoting agents (Roberfroid et al., 2010).

Currently, there is a high number of carbohydrates marketed as prebiotics, however only inulin and fructo-oligosaccharides (FOS), human milk oligosaccharides (HMOs), galacto-oligosaccharides (GOS), and lactulose can be considered as such, based on the scientific evidence obtained in human trials (Villamiel et al., 2014; Walton et al., 2012; Roberfroid et al., 2010; Falony et al., 2009). Considering structure-biological activity relationships, there is an overall interest in finding novel potentially prebiotic carbohydrates that may exhibit an enhanced bioactivity. Among novel molecules, isomalto-oligosaccharides, α-galactosides, gentio-oligosaccharides, xylo-oligosaccharides and oligosaccharides derived from lactulose have been considered as

potential prebiotics (Corzo et al., 2015; Cardelle-Cobas et al., 2008; Playne & Crittenden, 1996). Other promising candidate prebiotics are pectins and pectic oligosaccharides (POS) derived from pectin depolymerization process (Gullón et al., 2013).

3.3. Artichoke by-products as a source of functional ingredients

The valorization of industrial by-products to mitigate the negative environmental impacts of fruit and vegetable processing industries is an essential step in the transition toward a bioeconomy (Satari & Karimi, 2018; Banerjee et al., 2017). With this aim, green valorization schemes that lead to an integrated biorefinery platform have been introduced (Satari & Karimi, 2018). Plant-based food manufacturing sector is currently under expansion through development of novel products such as edible flowers, like traditionally globe artichoke (Gostin & Waisundara, 2019). Among the by-products originated by vegetable transforming industry, those generated by the artichoke (*Cynara scolymus*) processing are being studied as a possible source of biologically-active ingredients (Gostin & Waisundara, 2019; Terkmane et al., 2016). Artichoke is an edible vegetable widely consumed in the Mediterranean diet, harvesting in this area the 85% of world production (Llorach et al., 2002). Artichoke global market is dominated by Italy followed by Egypt and Spain, with an estimated annual harvest of 366 kt, 236 kt and 186 kt, respectively (Gostin & Waisundara, 2019). The industrial processing of artichoke produces a high amount of wastes (external bracts, leaves and stems), which is close to 60% of the harvested artichoke (Llorach et al., 2002). Although artichoke residues are generally used for animal feed or as manure (Machado et al., 2015), it can be used to extract bioactive compounds such as flavonoids and phenolic compounds (Sałata & Gruszecki, 2010; Llorach et al. 2002) and inositols (Ruiz-Aceituno et al., 2016) as well as to isolate fractions enriched in soluble fibre (Fissore et al., 2014).

In addition, artichokes synthesize and accumulate inulin (Gostin & Waisundara, 2019; Lattanzio et al., 2009; Schütz et al., 2006; López-Molina et al., 2005), which resist the small intestinal digestion due to the $\beta(2 \rightarrow 1)$ fructosyl-fructose linkages, but

can be fermented by colonic microflora (Gostin & Waisundara, 2019; Terkmane et al., 2016).

Biological properties of artichokes are related to their high levels of polyphenolic compounds and inulin (Gostin & Waisundara, 2019; Colantuono et al., 2018) and have been demonstrated in *in vitro* and *in vivo* studies including hepatoprotective and anticarcinogenic properties (Gostin & Waisundara, 2019; Ferracane et al., 2008). Other activities reported include potential anti-inflammatory, antioxidant, anti-genotoxic and anti-obesity effects (Colantuono et al., 2018; Kollia et al., 2017) as well as antimicrobial and anti-HIV activities (Dias et al., 2018; Lattanzio et al., 2009) and hypolipidemic properties due to the presence of inulin in synergy mainly with arabinans and arabinogalactans (Villanueva-Suárez et al., 2019).

Pectin is another functional ingredient that may be extracted from artichoke residues (Fissore et al., 2014). Recently, Santo Domingo et al. (2019) reported that artichoke by-products can be used to extract soluble dietary fiber fractions and different pectic polysaccharides that might be utilized as natural thickeners and gelling agents in food manufacturing.

3.4. Pectins

3.4.1. Structural characteristics

Pectins are a type of soluble fibre commonly present in vegetal cell wall (**Figure I1**) and are one of the most structurally complex families of polysaccharides in nature. Pectin exerts a great influence on plant growth, morphology, wall structure and porosity, signalling, development and plant defence. Pectin is formed by different cell wall polysaccharides (**Figure I2**) containing a linear chain of galacturonic acid (GalA) GalA-α(1,4)-GalA units called homogalacturonan (HG) which comprises approximately 65% of pectin. HG is partially methyl-esterified at the *O-6* carboxyl and may be O-acetylated at *O-2* or *O-3*. The most abundant pectin neutral sugars are xylose (Xyl), arabinose (Ara), rhamnose (Rha) and galactose (Gal). Thus some GalA units may be substituted at *O*-3 with β-D-Xyl, forming xylogalacturonan (XGA) which is more prevalent in reproductive tissues and may confer HG resistance to endo-

polygalacturonase enzymes produced by pathogen microorganisms. HG chains usually are formed by approximately one hundred GalA monomers in length although shorter HG chains might be interspersed between other structural domains. HG is the backbone that presents ramified chains of rhamnogalacturonan types I (RG-I) and II (RG-II) (Atmodjo et al. 2013; Mohnen, 2008).

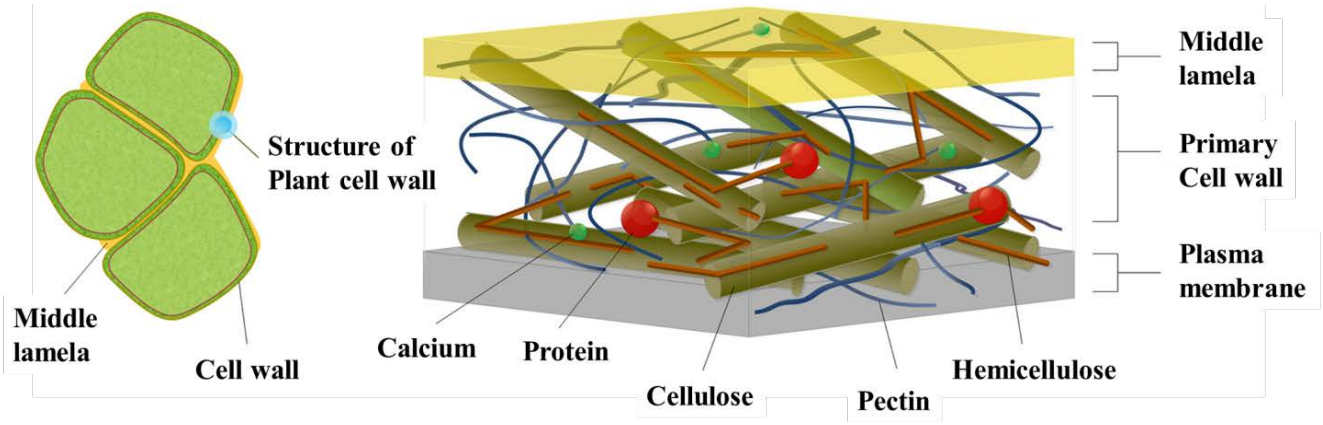


Figure I1. Structure of primary plant cell wall (adapted from Dranca & Oroian, 2018).

RG-I comprises 20-35% of pectin structure and is formed of sequences of disaccharide [-4)-α-D-Gal*p*A-(1,2)-α-L-Rha*p*-(1-]$_n$. In this domain, rhamnose may be substituted at *O*-4 with linear or branched oligosaccharides of arabinose (arabinans), galactose (galactans) or both (arabinogalactans) showing DPs up to 47 units. In addition, GalA units may be acetylated (Broxterman et al., 2017). The extent of branching depends on neutral sugar content so it may vary according to pectin source (Gullón et al., 2013; Caffall & Mohnen, 2009; Mohnen, 2008).

RG-II constitutes a minor and complex domain which comprises 10% of pectin structure. This domain consists of 12 different types of sugars in over 20 different linkages and is largely conserved across plant species (Atmodjo et al. 2013; Mohnen, 2008). Some monomers present in this domain are fucose (Fuc), apiose (Api), aceric acid (AceA), deoxy-D-manno-octulosonic acid (Kdo) and 2-keto-3-deoxy-d-lyxo-heptulosaric acid (Dha). Some complex chains described in the bibliography include the following dimers in their structure: [-Fuc-α(1,4)-GalA-], [-Rha-α(1,5)-KDO-], [-Ara-α(1,5)-Dha-], [-Rha-β(1,3)-Api-], [-Ara-β(1,2)-Rha-], [-Gal-β(1,2)-AceA-] (Pabst et al., 2013).

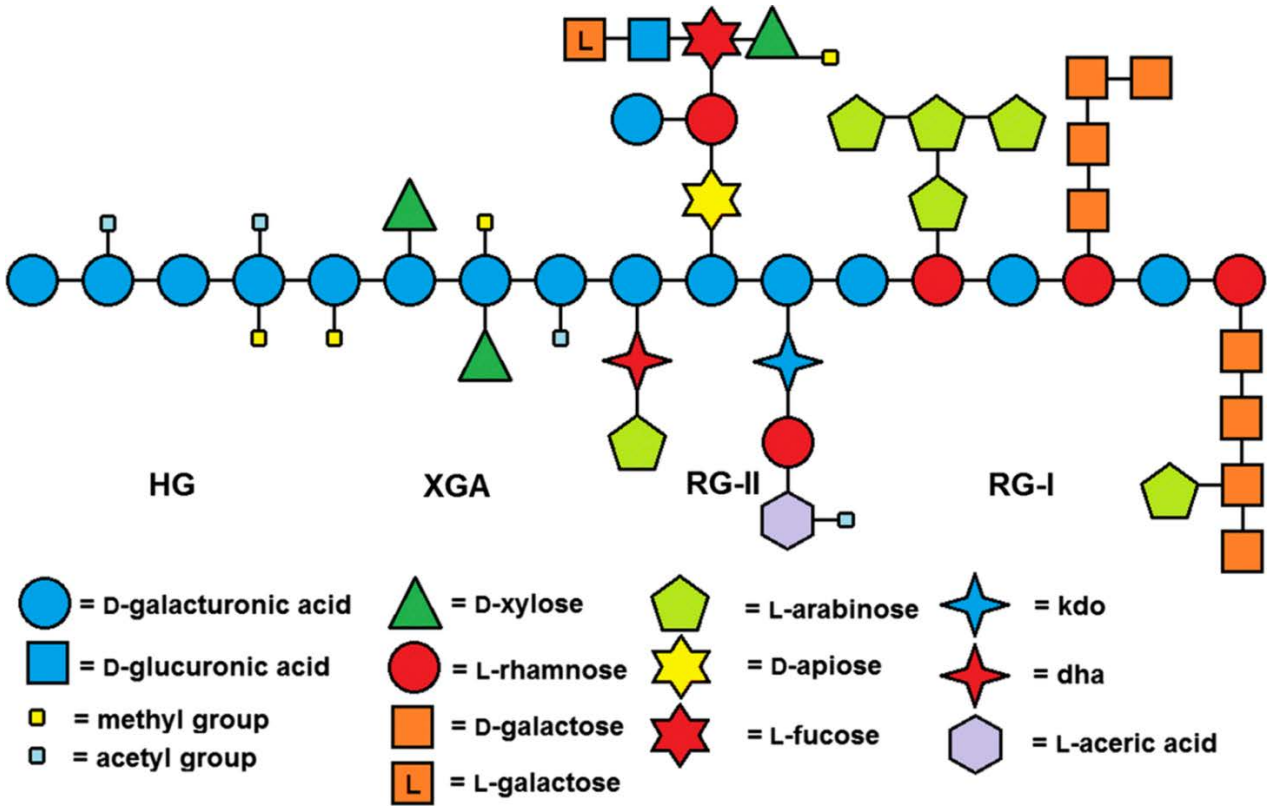


Figure I2. Schematic representation of pectin domains (adapted from Kuivanen et al., 2019). **HG**: homogalacturonan, **XGA**: Xylogalacturonan, **RG**: Rhamnogalacturonan. **Kdo**: Deoxy-D-manno-octulosonic acid, **Dha**: 2-keto-3-deoxy-d-lyxo-heptulosaric acid.

The abundance of each domain as well as the molecular weight (M_w) of pectin show a great variability between different pectin sources, ranging the later from 100 to near to 1000 kDa, depending on the pectin source and extraction method (Yang et al., 2018a,b). Other of the most relevant structural characteristics of pectin is the degree of methyl-esterification (DM), indicating the percentage of GalA esterified with methyl groups in conjunction with the total of this uronic acid content. Some pectin may show DM values up to 90% depending on the plant species, tissue and maturity. Pectin is classified according to their DM into high methyl-esterified pectin (HMP, DM values above 50%) and low methyl-esterified pectin (LMP, DM values below 50%). This parameter is strongly related to physical properties like viscosity and gelling capacity of pectin solutions (Muñoz-Almagro et al., 2018; Pagan et al., 2001).

3.4.2. Methods of extraction

Extraction process exerts great influence on pectin structure and may result in different technological/biological properties. There is no standard extraction technique (Gerschenson, 2017) although the procedures applied for commercial pectins are usually optimized to enhance the content of HG region which exert the useful functional gelling properties of pectin (Maxwell et al., 2012). Conventional pectin extraction consists the suspension of the raw material in aqueous acid solution; then, the mixture is filtered and the pectin precipitated using organic solvents such as ethanol (Marenda et al., 2019). With this aim, strong acids such as hydrochloric, nitric, and sulphuric acid and high temperature (70–100 ºC) has been traditionally used in the industry (Marić et al., 2018; Adetunji et al., 2017; Min et al., 2011). Pectin extraction employing acids is an efficient and economical process. However, it may be environmentally hazardous and could modify the structure of pectin. It should be noted that high temperatures may produce de-esterification and depolymerization of pectin chains (Marenda et al., 2019). Nowadays, pectin extraction using other weak acids and extracting food grade agents such as citrate has been reported (Marić et al., 2018; Muñoz-Almagro et al., 2018; Adetunji et al., 2017).

To overcome the limitations of chemical extraction, several alternative methodologies have been developed. Enzymatic extraction of pectin has been proposed as an environmentally friendly alternative to acidic extraction with the advantage of original pectin structure conservation (Marić et al., 2018). Other advantages of enzymatic process are its low equipment cost as well as low temperatures and use of low concentration buffer as solvent. This extraction method is performed using enzymes showing cellulase and hemicellulase activities which hydrolyse the cell wall increasing the cell permeability and released pectin molecules. The type of enzyme selected determines the yield of extraction and monomeric composition of pectin (Wikiera et al., 2015a; Dominiak et al., 2014). Commercial cellulase Celluclast® 1.5L (Celluclast) from *Trichoderma reesei* achieved complete pectin extraction from apple pomace with hardly any modification in its structure due to the wide range of activities present in this preparation: xylano-, cellulo-, and manno-lytic activities although can be highly time consuming (Wikiera et al., 2015a). It has been reported that enzymatically-extracted

pectin showed higher GalA contents, DM and degree of acetylation than pectin extracted with acids (Wikiera et al., 2015b). This fact may result in different technological and biological properties.

Other alternative techniques include microwave-assisted extraction, which operates at low temperatures for short times and achieves higher yields than conventional extraction. However, the main drawback of this technique is the high investment in equipment and maintenance needed for large-scale processes (Marenda et al., 2019; Roselló-Soto et al., 2016). High-pressure extraction shows several advantages including low temperatures, short extraction times and less use of solvents. Similarly, pectin extraction employing subcritical water requires short extraction times and less use of solvents although it operates at high temperatures (up to 170 °C) (Marenda et al., 2019; Wang et al., 2014a). Both high-pressure and subcritical water extraction lead to lower pectin yields compared to other methodologies. Ultrasound-assisted extraction is one promising and innovative technique that may overcome some of the limitations of the extraction methods described above. Power ultrasound (US) at low frequency (20-100 kHz) produce cavitation effect favoured assisted extraction of soluble compounds such as pectin, showing several advantages including high extraction yield, low energy consumption and short processing time. This methodology is based on the propagation of US waves involving several expansion and compression cycles that separates and joins the molecules alternately (Adetunji et al., 2017). In liquid media, the expansion cycles generate bubbles that grow and collapse in a process known as cavitation. Acoustic cavitation, produced by low frequency US, is the basis for extraction enhancement, generating microjets, shear forces and shockwaves. Cavitation creates points with very high temperatures (which increase solubility and diffusivity) and pressures (which facilitate penetration and mass transfer) at the interface between a liquid medium and solid matrix subjected to ultrasonication. Also, the microjetting and microstreaming effects attributed to acoustic cavitation cause a strong physical agitation at the interface between a liquid medium and solid matrix and a disintegration of solid materials and disruption of cell walls. This phenomenon involves an increase in the contact between the solvent and the cell contents, and enhances mass transfer of the cell contents from the material to the solvent, increasing the extractive power (Chandrapala et al., 2013; Soria et al., 2017). However, US extraction is a low selective methodology, extracting polysaccharides in general, often with a low GalA content (Dranca, &

Oroian, 2018). To overcome this drawback, US may be applied in combination with acidic or enzymatic treatments.

The combination of US and enzymatic treatments can provide potential benefits, shortening the process. In general, it has been observed that, despite the reduction of the enzyme activity, the outcome of combined process was positive (Nadar et al., 2018; Soria et al., 2017; Szabó & Csiszar, 2013). Numerous authors studied procedures for ultrasound-assisted enzymatic extraction (UAEE) of biomolecules obtaining positive results. Thus, Konwarh, et al. (2012) applied an UAEE procedure for lycopene extraction from tomato peel and they reported that the coupled system improved the extraction by 662, 225 and 150% times over the unaided, only enzymatic or only sonicated samples, respectively. Wang et al. (2014b) extracted arabinoxylan, founding that the extraction yield noticeably increased with US power from 50 W (6%) to 200 W (12%), while power above 300 W resulted in a lower yield (6%). Similar results were obtained by Liu et al. (2014) during UAEE of polysaccharides from *Lycium barbarum* obtaining under optimum conditions (20 min) a yield of 6.3%, which was almost similar to the yield of 6.8% obtained by enzyme-assisted extraction but much effective due to the extraction time needed (91 min) at the same conditions. Pardo-Rueda et al. (2015) extracted water-soluble carbohydrates from the sotol plant (*Dasylirion leiophyllum*) with different methods and observed that with UAEE treatment reach the highest carbohydrates content with a high fructan proportion (69%) compared with only thermal, pre-enzymatic or sonothermal treatment. Pu et al. (2015) extracted polysaccharides from rhizome of *Atractylodes macrocephala* and with an UAEE was obtained the higher extraction yield (13.2%) than without US (8.4%). Similarly, Xu et al. (2015) applied this combined technology to optimize the extraction of polysaccharides from black currant. The influence of US waves on the activity and stability of enzymes has been shown to be specific for each enzyme and dependent on sonication parameters (Rico-Rodríguez et al., 2018; Delgado-Povedano & de Castro, 2015). As it is showed, UAEE of pectin offers several advantages over traditional acid processes. These methodologies also conserve the original pectin structure that may have unique functional properties.

The majority of pectins used in the industry are extracted from citrus, apple and sugar beet by-products (Grassino et al., 2018; Adetunji, et al., 2017). However, alternative sources of vegetables could give rise to pectins with different structures and enhanced technological and biological properties are being studied. With this aim, non-

traditional sources of pectins have been studied such as by-products from cranberry, onion, garlic, banana, pumpkin, peach, rapeseed, papaya, fresh peas pomace, sunflower, bark of mango tree, red fruit pulps or sisal wastes (Adetunji et al., 2017; Santos et al., 2013). Interestingly, some alternative sources like tomato and papaya may show higher pectin contents (up to 49.8 - 83.5%) than traditional pectin sources (up to 20.9-33.6%) (Chan et al., 2017).

3.4.3. Technological properties

Pectins are widely used as valuable functional food ingredients due to their technological and biological properties. These properties are directly influenced by its structural characteristics such as monomeric composition, M_w, presence and distribution of side chains, DM and degree of acetylation, and charge distribution along their backbone (Herbstreith & Fox, 2018; Marić et al., 2018), and these structural parameters, are influenced by location and other environmental factors of crop and conditions of extraction (Fissore et al., 2014; Gullón et al., 2013; Canteri et al., 2012).

Nowadays, pectin technological properties determine its use as food ingredient in the industry. This polysaccharide is soluble in water at low concentrations (1-2%) and in other solvents like formamide, dimethylformamide or glycerine. Although pectin original structure shows a neutral charge, it may exhibit negative charge and acid pH when dissolved in water solutions due to GalA groups, depending on its DM, showing pH values between 2.8 and 3.4. Pectin solubility and viscosity depend on several factors including pH, temperature, DM, M_w and electrolyte concentration, increasing at solvent temperatures close to the boiling point (Pagan et al., 2001). A linear relationship between pectin concentration and the viscosity of the solution has been reported for pectin from apple, cacao husk and citrus peels (Chan et al., 2017).

The pectin DM also exerts a great influence on technological properties, especially on gelling capacity. So HMPs show high viscosities when present high M_w and degree of branching, forming gel at acid pH (2.8 - 3.5) in solutions containing 60-70% soluble solids. In contrast, LMPs show high viscosities in the presence of divalent cations like calcium (25 – 100 mg g^{-1} pectin) that interact with free carbonyl groups of GalA leading to formation of complex (Muñoz-Almagro et al., 2018; Pagan et al.,

2001). It should be noted that LMP gels are formed in a wider range of pH (1.0 – 7.0) and soluble solid content (0 – 80%) than those from HMPs. With respect to emulsification, RG-I and RG-II domains may confer pectin an emulsifying capacity due to their potential interaction with proteins (Funami et al., 2011). This property is heavily influenced by the degree of acetyl esterification and neutral sugar content of side chains (Leroux et al., 2003).

Pectin is widely used food additive (E-440) because of these technological properties (gelling agent, stabiliser and fat replacer) (Dominiak et al., 2014). Traditionally, HMP are employed as emulsifying and gelling agents in acid matrices with high sugar contents such as jam and yogurt, however currently, LMP are being used for their gelling properties, since they may be incorporated in a wider variety of products with lower sugar content. Several studies reported the applications of pectin in dairy product formulation, one of the main segments in functional food development being yogurt one of the most popular fermented milk consumed worldwide and a good vehicle to delivery functional ingredients (Demirkol & Tarakci, 2018). It has been demonstrated that citrus pectin enhances rheological quality of yogurt and leads to a better proliferation of *Streptococcus thermophilus* and *Lactobacillus bulgaricus* during the period of fermentation (Arioui et al., 2017). It should be noted that HMP-supplementation produces significantly higher thickness than other hydrocolloids (Gallardo-Escamilla et al., 2007). Among the potential applications of LMP, it was observed an increase in oral consistency in goat milk yogurts (Bruzantin et al., 2016). Other studies found that pectin enhances flavour in probiotic yogurt during storage (Karaca, 2013) and whey protein-HMP complexes could be used as fat replacers in yogurts (Krzeminski et al., 2014). While most of these studies focus on the modification of technological and sensory properties of pectin-supplemented foods, few studies deal with the bio-functional properties of pectin when incorporated in these matrices.

3.4.4. Biological properties

Biological properties of pectin have gained great attention. Pectin is classified as soluble fibre because it cannot be digested in the gastrointestinal tract. However, it can be selectively fermented by the microbiota of colon so it can be considered as potential

prebiotic. This effect of pectin has been demonstrated in several studies where pectin stimulated the growth of *Bifidobacterium* and *Lactobacillus* population and inhibited the growth of *Clostridium* (Gullón et al., 2013). Ferreira-Lazarte et al. (2018) reported a significant increase of *Bifidobacterium* population during *in vitro* batch fermentation of citrus, sunflower and artichoke pectin using faecal inocula. It was observed that lower M_w values lead to an increase in pectin ability to promote bifidobacteria growth. A significant increase was also observed in *Bacteroides* and *Prevotella* populations. On the other hand, it has been demonstrated that LMP is more easily fermented than HMP (Olano-Martin et al., 2002). It should be noted that *Bacteroides* species are major carbohydrate metabolising organisms in the gut and have the capacity to degrade diverse plant polysaccharides. Similarly, an *in vitro* study using a dynamic gastrointestinal simulator (Ferreira-Lazarte et al., 2019) corroborated the high indigestibility of citrus pectin as well as its ability to stimulate the growth of beneficial bacteria such *Bifidobacteria*, *Bacteroides* and *Faecalibacterium prausnitziiand*. Short chain fatty acids (SCFA) formation was also reported due to fermentation of pectin. Other beneficial bacteria like *Bacteroides thetaiotaomicron* may ferment pectin producing SCFAs as acetate, propionate and butyrate which contribute to normobiosis (Gerschenson, 2017). Interestingly, *F. prausnitzii*. is considered as a marker for intestinal health, exhibiting anti-inflammatory effects in the gut (Larsen et al., 2019; Walters et al., 2014). The populations of some bacterial associated with human health can be modulated by specific pectins. In this sense, growth of *F. prausnitzii* is stimulated by HMP while LMP decrease the populations of *Prevotella copri* associated with rheumatoid arthritis (Larsen et al., 2019; Pianta et al., 2017; Pedersen et al., 2016). So the application of specific pectin structures and conformations may result in a higher formation of SCFAs from proximal to distal colon regions (Gerschenson, 2017).

SCFAs are also associated to numerous benefits on human health like reduction of risk of gastrointestinal and cardiovascular diseases (Gullón et al., 2013). With regard to its anti-inflammatory activity, Pacheco et al. (2018) reported that commercial citrus pectin and industrial products rich in pectin (animal feed and fresh orange residue) ameliorated some IBD symptoms in mice and gave rise to a lower expression of pro-inflammatory cytokines, intercellular adhesion molecules ICAM I, iNOS enzyme, and a higher expression of protective chemokines MUC 3, occludin and ZO-1. An anti-carcinogenic activity of pectin has been also reported; pectin may reduce colon cancer risk as pectic polysaccharides and modified pectins (smaller pectin fragments with M_w

comprized around 100 and 7 kDa), exhibited an antiproliferative activity on several cancer cell lines (Gerschenson, 2017). Arabinan-rich pectic polysaccharide exerts an immunomodulatory activity by stimulating the expression of several lymphocyte activation markers (Tingirikari, 2018) while RG-II promotes a tumour-specific cell-mediated immune response in mice (Tingirikari, 2018; Park et al., 2015). Pectin may also form complexes with bile acids and glucose reducing blood cholesterol levels and/or intestinal absorption of glucose (Morris et al., 2013). Furthermore, pectin showing GalA contents below 75% may have an immunomodulatory activity due to the large presence of RG-I chains (Popov & Ovodov, 2013a).

Considering these beneficial properties, pectin could be incorporated in functional foods as a biologically active ingredient although it shows relatively low solubility and high viscosity, so pectin can be added in doses up to 1%, and this concentration is very low to favoured the biological activity of polysaccharide. To fit different technological/functional purposes, these pectins could be used as substrate to obtain modified pectin and POS following chemical or enzymatic methods.

3.5. Pectic oligosaccharides (POS)

Depolymerization of polysaccharides is the most reliable and widely used process for oligosaccharide production on a large scale achieving high yields (Moreno et al., 2017). In accordance with this consideration, novel oligosaccharides with improved functional properties can be obtained using plant cell wall polysaccharides as raw material (da Moura et al., 2015).

POS are produced from depolymerization of raw materials or purified pectins usually extracted from apple, citrus peel and sugar beet by-products, using enzymes, chemical reagents (acids) or combined methods (Gullón et al., 2013; Martínez et al., 2009a,b). Also other substrates such as onion skins (Baldassarre et al., 2018), rapeseed cake (Cobs-Rosas et al., 2015) have been used to obtain POS. Due to complexity of pectin, resulting POS can have varied chemical structure which will have influence in their bioactive properties. POS obtainment is usually carried out in two different steps: pectin extraction from the raw material and pectin hydrolysis. To optimise this process,

an alternative method using a combination of enzymes to produce POS from the raw material has been developed (Babbar et al., 2017).

Chemical depolymerization of pectin involves partial hydrolysis of pectin chains through acid treatments at high temperature/pressure conditions (Berlin et al., 2007). Chemical treatments using acid solutions produced RhaGalAOS from sugar beet pulp, while GalAOS with DP up to 10 have been obtained following hydrothermal treatments (Gullón et al., 2013). These processes show some disadvantages including the high cost of high pressure equipment. In addition, acid treatments are environmentally hazardous and may modify the original structure of pectin/POS chains through de-esterification.

3.5.1. Enzymatic obtainment of POS

Enzymatic hydrolysis of pectins offers several advantages, such as high regio- and stereoselectivity, obtaining structurally different POS mixtures (Babbar et al., 2016a,b; da Moura et al., 2015). Most food-grade POS studied are produced by pectinases, a complex group of enzymes that degrade pectic substances present in plant tissues (**Figure I3**). In general, these enzymes are classified into de-esterificating and depolymerizing enzymes based on the enzymatic activities (Gummadi & Panda, 2003). Polygalacturonase (PG; E.C. 3.2.1.5) hydrolyses α-(1,4) linkages between D-GalA present in the HG. Endo-PG randomly hydrolyses the HG while exo-PG involve the sequential hydrolysis of α-(1,4) linkages at the reducing end of the HG chain. Therefore, exo-PG leads to larger pectin chains and GalA monomers released while endo-PG produces several types of POS showing a wide range of M_w. For this reason, endo-PGs are employed to achieve high yields of POS, up to 95%, with low M_w (Gullón et al., 2013; Olano-Martin et al., 2001). On the other hand, pectin lyase (PL; E.C. 4.2.2.10) hydrolyses α-(1,4) linkages by trans-elimination resulting in GalA units with an unsaturated bond between C4 and C5 at the non-reducing end of the GalA formed (Tapre & Jain, 2014). Finally, pectin esterase (PE; E.C. 3.1.1.11) catalyses deesterification reactions of methyl-ester groups, not releasing POS. PE acts preferentially on a methyl-ester group of galacturonate unit next to a non-esterified GalA unit (Tapre & Jain, 2014). To date, most food-grade POS studied have been produced by commercial enzymes, mostly pectinases, usually produced by *Aspergillus*

spp. (Ma et al., 2018; Combo et al., 2012; Delattre et al., 2008) which synthesized a battery of enzymes to hydrolyse cell wall-polysaccharides (de Vries & Visser, 2001).

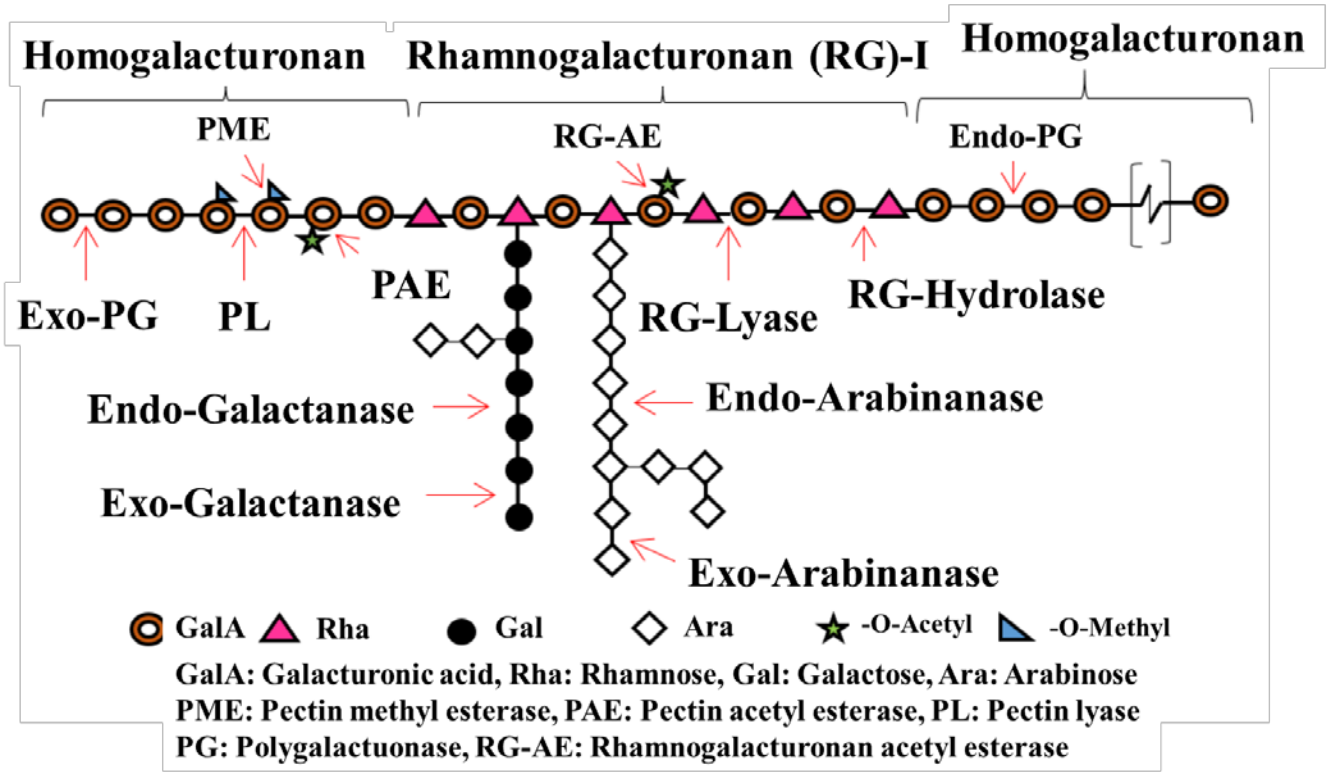


Figure I3. Examples of different homogalacturonan and rhamnogalacturonan I degrading enzymes (adapted from Álvarez, 2015).

However, several studies reported POS formation using enzymes from other microorganisms such as *Bacillus* spp. (Yu & Xu, 2018, Rehman et al., 2015; Uzuner & Cekmecelioglu, 2015), *Saccharomyces cerevisiae* (Poondla et al., 2015) and *Penicillium viridicatum* (Silva et al., 2002). Other enzymes that catalyse the hydrolysis of pectin chains involve rhamnogalacturonases, galactanases and arabinases that lead rhamno-oligosaccharides with low DP (Bonnin et al., 2002). Enzymatic treatments using four different types of enzymes (PE, endo-PG, endoarabinase and endogalactanase) lead to POS formation with a DP up to 18, with predominance of low M_w GalA oligomers (Ralet et al., 2008).

3.5.2. Structure of POS

According to different criteria, POS are formed by 2 – 50 monomeric units or can show M_w values up to 7 kDa (Babbar et al., 2017; Li et al., 2016; Bonnin et al., 2014). In contrast, larger fragments comprized between 7 and 100 kDa could be considered as modified pectin (Ognyanov et al., 2016; Dominiak et al., 2014; Ramasamy, 2014; Westereng et al., 2009).

These POS can present very different chemical structures depending on the depolymerization method, reaction conditions and substrate used (da Moura et al., 2015). In general, POS mixtures derived from RG-I are formed of various types of oligomers containing up to 14 different monomers bonded by α and β linkages, i.e. oligogalacturonides (GalAOS), galacto-oligosaccharides (GalOS), arabino-oligosaccharides (AraOS), rhamnogalacturonides (RhaGalAOS), xylo-oligogalacturonides (XylGalAOS) and arabinogalacto-oligosaccharides (AraGalOS) (Gullón et al., 2013). The monomeric composition of these structures is determined by the original domain of pectin (HG, XGA, RG-I, RG-II). In general, HG [-4)-α-GalA-(1,4)-α-GalA-(1-] and XGA chains (β-D-Xyl-(1,3)-[GalA-(1-]) may be both methyl- or acetyl-esterified. With regard to RG-I-derived POS, GalA units may be acetyl esterified but not methyl-esterified has been found. In addition, galactose units are only bonded to rhamnose by β(1,4) linkages for galactans and arabinogalactans. Therefore, POS dimer units formed from RG-I can be (acetylated or not) GalA-α(1,2)-Rha, Rha-α(1,4)-GalA (acetylated or not), Gal-β(1,4)-Rha and Ara-α(1,4)-Rha, for arabinans (**Figure I4**) according to the criteria established by several authors (Atmodjo et al. 2013; Mohnen, 2008). However, it must be considered that these are the main but not the only estructures present in pectin, since other monosaccharides such as apiose, fucose or glucuronic acid and other linkages as α(1,3) and α(1,5) may be present. Larger POS structures are formed mainly by several combinations of the previous units.

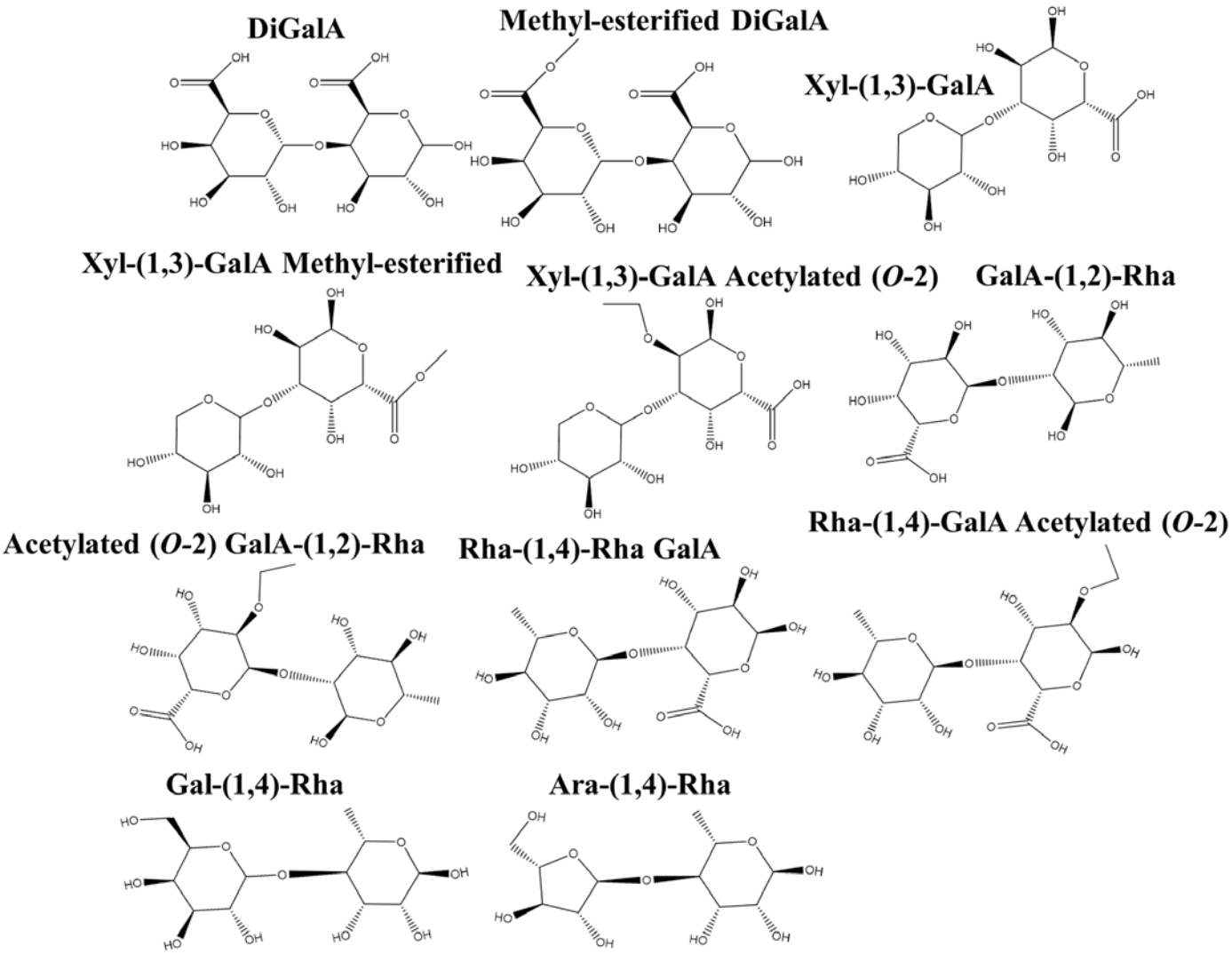


Figure I4. Most probable dimers present in POS structures released from homogalacturonan, xylogalacturonan and rhamnogalacturonan-I domains from hydrolysis of pectin. **GalA**: Galacturonic acid, **Xyl**: Xylose, **Rha**: Rhamnose, **Ara**: Arabinose, **Gal**: Galactose.

3.5.3. Biological properties of POS

POS from different sources show an enhanced bioactivity compared to the original pectin chains (Gómez et al., 2016; Li et al., 2016; Olano-Martin, et al., 2002). POS are non-digestible oligosaccharides because of the conformation of the anomeric carbon (C1) in their monomeric units, which is not recognized by human digestive enzymes (Roberfroid & Slavin, 2000). POS have been proposed as a new class of prebiotics due to their bifidogenic effect, resulting in SCFAs formation and a decrease in

the pH. As a consequence, these oligosaccharides inhibit the growth of pathogen bacteria reducing the risk of infection. *In vitro* and *in vivo* studies demonstrated that acidic POS are neither cytotoxic nor mutagenic, the first requirement for their use in all foods, i.e. infant formula (Gullón et al., 2013). Moreover, POS are non-cariogenic sugars. In general, these properties make them attractive to functional food development (Rubio-Arraez et al., 2016; Mussatto & Mancilha, 2007).

Interestingly, neutral oligosaccharides showing a low DP exert a bifidogenic effect during the first hours of colonic fermentation (Bonnin et al., 2014) while acid oligosaccharides do not exert this effect. In this sense, other studies (Van Laere et al., 2000) reported the selective fermentation of RhaGalAOS by *Bacteroidetes* spp. whereas *Clostridium* spp. did not utilize these compounds. It was also observed that AraGalOS and AraOS but not RhaGalAOS or GalAOS were fermented selectively by *Bifidobacterium* spp. Olano-Martin et al. (2002) found that POS showing low DM were fermented by *Bifidobacterium angulatum*, *Bifidobacterium infantis* and *Bifidobacterium adolescentis*. In general, bifidobacteria are able to utilize a wide variety of POS with DP ranging from 3 to 7 which were not fermented by *Bacteroides* spp. and clostridia (Mandalari et al., 2007). In addition, small arabinans (DP 2 - 3) stimulated *Lactobacillus* growth and similar oligosaccharides with DP 2 - 5 have shown to have promising prebiotic activity for *Lactobacillus brevis*, *Bifidobacterium longum* and *Bacteroides fragilis* (Prandi et al., 2018).

Colonic fermentation patterns of POS with different structures have been also established using faecal inocula. Al-Tamimi et al., (2006) reported that low M_w AraOS were more selectively fermented by *Bifidobacterium* spp. than larger oligosaccharides with similar structures. The influence of M_w on its *in vitro* fermentability using human faecal inocula was also confirmed by Holck et al. (2011) in AraOS mixtures (DP 2 – 14) obtained from sugar beet pulp. This parameter (M_w) was even more relevant than the degree of feruroyl substitution. On the other hand, few studies evaluating POS effects in humans have been carried out. The administration of infant formula supplemented with acidic POS led to an increase in bifidobacteria and lactobacilli populations (Magne et al., 2008; Fanaro et al., 2005) while enteral delivery of both neutral and acidic POS resulted in a lower incidence of infections with endogenous bacteria (Westerbeek et al., 2009). Moreover, it has been demonstrated that mixtures of POS with other prebiotics as GOS and FOS had positive immune effects, because the addition of the former facilitates the required oligosaccharide diversity and their acidity enabling them to

interact with surfaces and to prevent the adhesion of pathogens on the intestinal epithelium (Jeurink, et al. 2013).

Other important beneficial effect of POS is the prevention the adhesion of pathogen microorganisms including *Campylobacter jejuni* to epithelial cell receptors. With regard to systemic effects of POS, it has been reported that POS with M_w of 10 kDa or lower shown an improved binding capacity to galectin-3 receptor, a protein whose expression is deregulated and overexpressed in cancer cells, so POS are being studied for the maintenance therapy of patients with lymphocytic leukaemia relapse (Karboune & Khodaei, 2016). In the same line, POS induces apoptosis of human colonic adenocarcinoma cells, also citrus POS exerted *in vitro* antitumor activity which was heavily influenced by M_w, GalA content and monomeric composition (Li et al., 2019a). Similarly, POS from hawthorn fruit showed a significantly inhibitory effect on the liver inflammation *in vivo* (Li et al., 2019b). Other beneficial activities studied are immunomodulatory activity, cardiovascular protection, reduction in blood sugar levels, and contribute to detoxifying heavy metals (Babbar et al., 2017; Gerschenson, 2017; Bonnin et al., 2014; Holck et al., 2014; Gullón et al., 2013).

3.6. Characterization of pectin and POS

Bioactive carbohydrates, both naturally occurring and synthesized, consists of complex mixtures which contain compounds with different DP, glycosidic linkages or monosaccharide composition. Considering that the bioactivity of a biomolecule is determined by its structure, fractionation and purification procedures are required for oligosaccharide characterization and bioactivity assessment, therefore new methods to purify specific POS structures are being developed (Holck et al., 2014). POS purification can be carried out by ultrafiltration (Gullón et al., 2013) while ion exchange chromatography and gel filtration chromatography are used for POS fractionation. Ultrafiltration and nanofiltration have been used to fractionate POS mixtures from fruit of *Actinidia arguta* with different M_w in order to further study the relationships between their structural parameters and their prebiotic activity (Zhu et al., 2019). Gel filtration chromatography has been used to purify and fractionate oligosaccharides from polysaccharide potato peels (DP up to 8) (Jeddou et al., 2018) and to isolate individual

potato galactan oligomers (DP ranging from 3 to 15) to study their prebiotic effect on *Lactobacillus* strains (Zheng et al., 2018).

3.6.1. Analytical techniques

As explained, POS mixtures can be formed of various types of oligomers and different analytical techniques can be employed to identify and characterise these compounds. Chromatographic techniques as high performance size-exclusion chromatography with refractive index detector (HPSEC-RID) and with evaporative light scattering detector (HPSEC-ELSD) (Muñoz-Almagro et al., 2018) determine complete M_w profiles of pectin and POS. It should be noted that HPSEC technique presents several limitations: M_w calculation might not be accurate especially for larger pectin and POS chains and specific structures of individual compounds present in the reaction mixtures cannot be determined. When characterising complex POS mixtures, it should be considered that discriminating between the two types of structures, neutral and acidic oligosaccharides, is especially relevant, highlighting to the presence of GalA, the main monomer of pectin. With this aim, anion-exchange chromatography coupled with pulsed amperometric detector (HPAEC-PAD) is the preferred option for quantifying both neutral and acidic species with different DP. HPAEC-PAD has been used to elucidate arabinoxylan structures in different wheat varieties (Ordaz-Ortiz et al., 2005) for arabinan and galactan oligosaccharide profiling (Wefers & Bunzel, 2016) and galactan oligomers with DP up to 15 (Zheng et al., 2018). This technique has been also applied to characterize small POS structures such as acidic and neutral POS, AraOS from sugar beet arabinan with DP comprized between 2 and 8 (Gómez et al., 2013; Al-Tamimi et al., 2006), analysis of GalAOS (DP 1–9) from polygalacturonic acid hydrolysates (Combo et al., 2012), the quantification of POS formation (DP 1-3) from sugar beet pectin (Combo et al., 2013), characterization of AraOS (DP 3–21), GalOS (DP 5–12), and OGalA (DP 2–12, with variable DM) from orange peel wastes (Gómez et al., 2014) and GalAOS mixtures in *Arabidopsis* with a distribution of DP between 8 and 20 (Pontiggia et al., 2015). Other examples include, the determination of POS from onion skins showing DP up to 9 (Babbar et al., 2016a) and identification of POS

fractions (GalAOS, AraOS and GalOS) with different Mw ranging from 1 to 3 kDa (Li et al., 2016).

Other most advanced techniques like hydrophilic interaction liquid chromatography coupled to mass spectrometry with electrospray ionization (HILIC-ESI-MS) (Leijdekkers et al., 2015) are used to deepen structural characterization of POS chains and to elucidation of individual POS species; several POS structures obtained from sugar beet pulp were determined using HILIC-ESI-MS: AraOS (DP 2 – 10) and GalA oligomers (DP 3-7) showing two methyl-esterifications and/or one acetyl esterification (Prandi et al., 2018). AraOS, RhaAraOS and RhaGalAraOS with DP ranging from 5 to 49 have been also determined (Doco et al., 2015). Simlarly, GalAOS and RhaGalAOS from bamboo shoots and potato have been elucidated by NMR and HPLC-ESI-MS (Ishii et al., 2002). Other advanced MS-based techniques like MALDI-TOF-MS have been used for a comprehensive characterization of ginseng RG-I and RG-II domains including methyl-ether and acetyl group sites in side chains (Sun et al., 2019) and to elucidate potato peel oligosaccharides with DP ≤ 20 (Jeddou et al., 2018). However, gas chromatography-mass spectrometry with electronic impact (GC-EI-MS) can be of great interest in small molecule characterization because of its robustness, reproducibility and relatively simplicity. GC-EI-MS fragmentation process for metabolite and small molecule identification has been widely studied (Allen et al., 2016; Meringer & Schymanski, 2013). This technique generates massive amounts of high-dimensional data.

In general, MS allows determining structural patterns of complex molecules and have become the analytical method of choice in metabolomics research. MS spectra need to be interpreted to extract valuable chemical information about the structures under study (Yi et al., 2016; Boccard & Rudaz, 2014). The identification of unknown compounds is the main bottleneck, so several computational tools assisting structure elucidation have been developed.

3.6.2. Applications of machine learning in metabolomics and food science

Several computational methods to assist MS data analysis and metabolite identification have been developed. The most widely used strategy involves reference mass spectral libraries, while other advanced methods to elucidate novel molecules, which may not be included in these libraries, are based on the use of machine learning (Bitchagno & Tanemossu, 2019; Nguyen et al., 2018). Machine learning consists of designing efficient and accurate prediction algorithms (Mohri et al., 2018) by programming computers to optimise a performance criterion using example data. These algorithms are defined up to some parameters and the learning process involves the execution of a computer program to optimise these parameters using training data. Machine learning models can predict new data, descript data to acquire knowledge from data, or both (Alpaydin, 2014). These models can be classified as i) supervized and ii) unsupervized methods. Supervized learning supports *a priori* known data structures to train patterns and rules to predict new data, so the algorithm is used to learn the function from the input (X) to the output (Y) variables. Supervized methods are grouped into regression and classification models. In classification tasks, the output is a categorical variable while in regression problems the output is a continuous variable. By contrast, unsupervized learning establishes unknown similarities and differences in the input data to model the underlying structure or distribution in the data in order to learn more about the data.

3.6.2.1. *Different types of models*

There are several families of algorithms that may be used to study chemical and biological data. When interpreting these models, it should be considered that each algorithm is computed in a different way leading to different results. These models may show a better/worse performance depending on the experimental data and different types of algorithms can be compared for different applications.

Some simple machine learning models are:

Generalized linear model (GLM). GLM consists in a linear regression model that could be generalized (i.e. the response variable may follow different distributions than Normal distribution).

Naïve Bayes classifier (NB). NB model is based on the Bayes Theorem that assumes that the presence of a particular feature (i.e. studied variable) in a class (i.e. type of sample studied) is unrelated to the presence of any other feature and calculates the probability for each class. The class with the highest probability is the outcome of prediction.

On the other hand, more advanced machine learning models include:

Random forest (RF). RF consists in multiple decision trees where each tree is built on a random subsample from the original data. Then, RF averages the predictions from all trees to get more accurate predictions than the individual predictions of each tree. It has been suggested that RF is a more robust model than other machine learning algorithms (Liaw et al., 2002) leading to a higher predictive power.

Boosted logistic regression (BLR). BLR is an ensemble of different simple logistic regression models, each one slightly different, that are used to build a powerful and complex model. The maximal number of simple models that are used is defined as iterations.

Artificial neural networks (ANN). ANNs are one of the most popular families of machine learning models used for pattern-recognition tasks. Among different ANN models, multilayer perceptron (MLP) is the most common kind of ANN. MLP allows modelling complex and highly non-linear processes and is formed by an input layer (i.e. studied variables, *m/z* ions), an output layer (i.e. sample type, more relevant *m/z* ions) and several neurons organized in hidden layers, where each neuron in a layer is connected with each neuron in the next layer through a weighted connection (**Figure I5**). In addition, a mathematical transformation (activation function) is applied to the input layer to determine whether the information that the neuron is receiving is relevant or not.

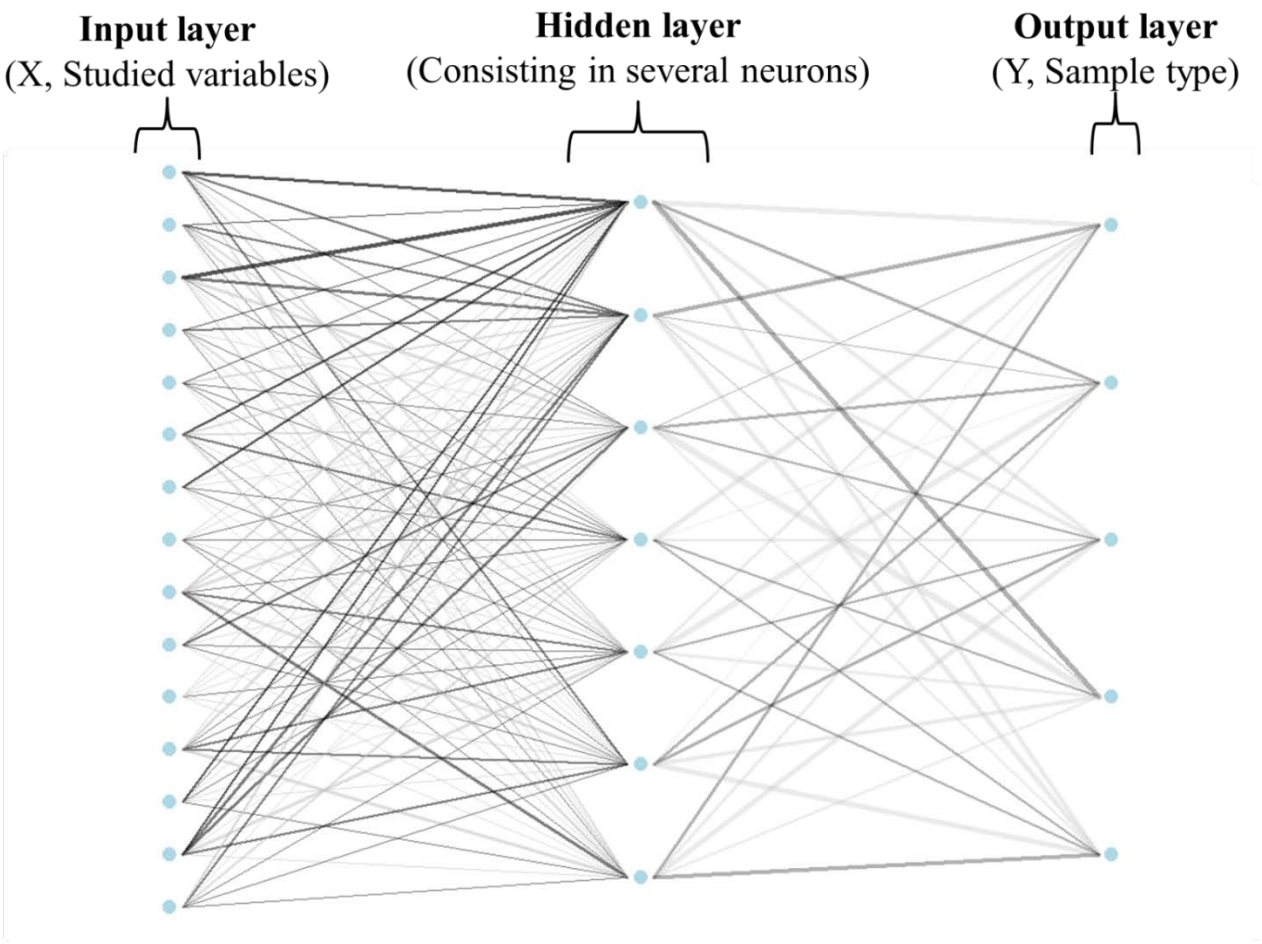


Figure I5. Multilayer perceptron (MLP) architecture consisting in one input layer and one output layer connected by several neurons (hidden layer). Weights can be positive (black) or negative (grey), and thickness is in proportion to its importance.

Convolutional neural networks (CNN). CNNs are a complex type of ANN used to classify images. This model simulates the mechanism of neurons in the visual cortex to capture the spatial and temporal dependencies in an image through the application of specific filters. CNN model consists in several layers called convolution layers, pooling layer, flatten layer and output layer (**Figure I6**). First convolution layer captures the low-level features of the images, while second layer captures the high-level features as well. Pooling layer is used for dimensionality reduction and extracting dominant features, while flatten layer (fully-connected layer) extracts non-linear combinations of the high-level features. As a result, CNN is able to make accurate predictions from image data (output layer).

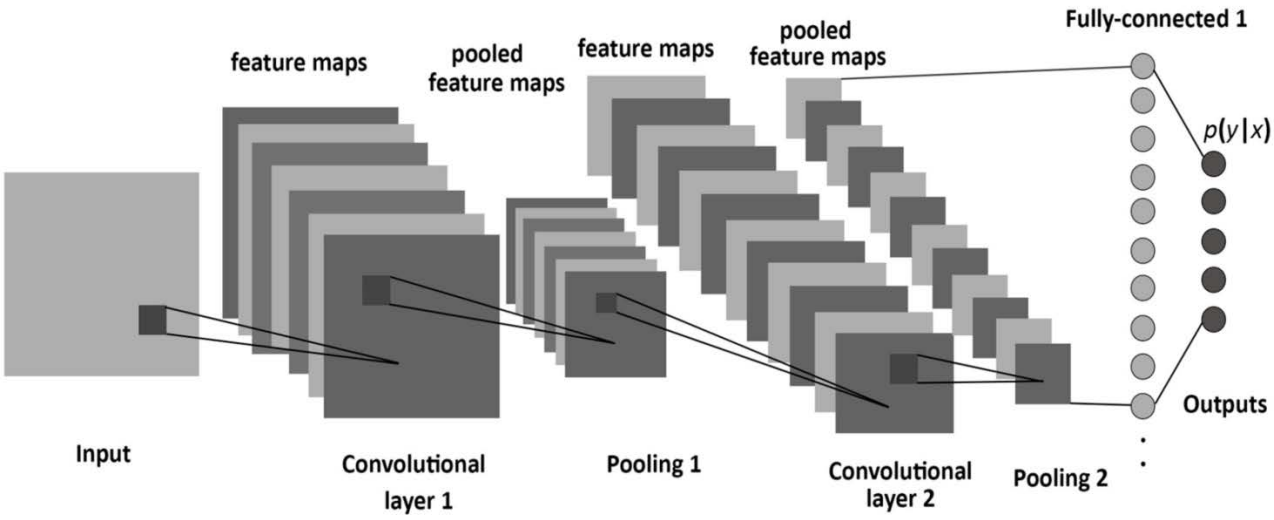


Figure I6. Convolutional neural network (CNN) architecture (Albelwi & Mahmood, 2017).

3.6.2.2. *Model applications to classify metabolomics and food science data*

Machine learning models can be used to integrate biology and medicine data which is often high dimensional, sparse, incomplete, biased and noisy to generate new knowledge from the existing data (Zitnik et al., 2019). These data modelling tools find reproducible patterns in samples to discover valuable information on biological events and chemical/structural properties (Yi et al., 2016; Käll et al., 2007). A great number of algorithms to determine classification patterns among samples have been developed (Kotsiantis et al., 2007) and applications of these models have been described to classify general metabolomics and food science data (Lim et al., 2017; Lin et al., 2014; Uarrota et al., 2014).

Omics applications. Applications in proteomic data profiling, classification and biomarker discovery (Gertheiss & Tutz, 2009), tumour classification (Behrmann et al., 2017) as well as metabolomic profiling of single cells (Liu et al., 2019a) and quantitative structure-property relationships of different molecules have been described (Rojas et al., 2019; Yu et al., 2019).

Food science applications. Several studies employ machine learning: for product characterization and classification milk samples under different storage conditions (Fabris et al., 2010), optimization of ultrasonic extraction of phenolic compounds from

wine lees (Tao et al., 2014), formulation of microparticles for polyphenol delivery (Belščak-Cvitanović et al., 2015), detection of adulterations (Lim et al., 2017), automatic specimen identification of crustaceans (Rossel & Arbizu, 2018), modelling of starch gelatinization (Tao et al., 2018), and early detection of *Aspergillus* spp. contamination in rice (Gu et al., 2019). In our research group, machine learning algorithms have been used to classify prebiotic-supplemented infant formulas according to their protein and carbohydrate sources (Sabater et al., 2018b). Recent applications of specific models described in the previous section (1.6.2.1.) have been also reported:

RF models in food science. Uses of RF for the automatic detection of adulterations in primrose oil and ground nutmeg (de Santana et al., 2019), prediction of *Listeria* spp. prevalence in pastured poultry farms (Golden et al., 2019), assessment of food quality and authenticity in cheese, coffee and several fruits (Jiménez-Carvelo et al., 2019) and qualitative identification of tea quality grades (Xu et al., 2019) have been reported.

ANN models in food science. ANNs practical applications involve classification studies of cow milk from different origins (Behkami et al., 2019), prediction of S-ovalbumin content in stored eggs (Fu et al., 2019), modelling of limonene release from amylose nanocarriers (Ganje et al., 2019), rapid determination of aflatoxin B-1 concentration in soybean oil (Liu et al., 2019b) and prediction of sensory quality of garlic (Liu et al., 2019c).

CNN models in chemistry and food science. Complex image-recognition models like CNN have been applied for automated targeted analysis of volatile organic compounds (Skarysz et al., 2018), recognition of several dish images (Im Cho et al., 2019), image classification of damaged and unusable potatoes in food manufacturing process (Jagtap et al., 2019), rice grain yield estimation in crop fields (Yang et al., 2019) and classification of flour, bean, corn, rice, and potato powders (You et al., 2019).

3.6.2.3. *Model applications to interpret MS spectra*

As explained, MS spectral libraries are the most widely used method for compound identification. Experimental data of unknown molecules are compared to an existing metabolite database matching a set of candidate molecules ranked according to spectral similarities (Blaženović et al., 2018; Nguyen et al., 2018). This method can be useful

when reference spectrum of the targeted compound is included in the database. Some popular databases are The Human Metabolome Database (HMDB) and Metabolite and Chemical Entity Database (METLIN) for metabolite profiling, as well as the National Institute of Standards and Technology mass spectra library (NIST) for GC-MS spectra matching (Acunha et al., 2018). However, it is not suitable for studying novel compounds which might not be previously described in the bibliography. To overcome the limitations of existing libraries, different computational approaches have been proposed. To interpret the results generated by advanced instruments like MS/MS spectra, specific techniques such as fragmentation trees might be employed (Böcke r& Rasche, 2008). These fragmentation trees represent the relationships between precursor ions of analytes (with their molecular formulas) and product ions generated in the MS/MS fragmentation reactions (Nguyen et al., 2018), and can be used to cluster molecules and to establish structural similarities among metabolites (Rojas-Cherto et al., 2012).

On the other hand, *in silico* fragmentation methods can be employed to simulate the fragmentation mechanisms of a wider variety of MS systems. This approach predicts *m/z* fragments and their abundance that may occur in MS spectra given the molecular structure of one compound, and may help in the deduction of possible structures of unknown metabolites (Nguyen et al., 2018; Allard et al., 2016; Ruttkies et al., 2016). By employing these methods, it is possible to identify MS and MS/MS spectra when the reference spectrum is not available. Most of *in silico* fragmentation software developed generate a MS spectrum following a set of general fragmentation processes previously extracted from existing data sets of elucidated MS/MS spectra, or calculate a series of substructures from a candidate molecule and then finds the most likely subset of the substructures that best matches the query spectrum giving a similarity score (Bitchagno & Tanemossu, 2019; Nguyen et al., 2018). Some commercial rule-based software used in the bibliography includes Thermo Scientific Mass Frontier, ACD/MS Fragmenter, MOLGEN-MSF (Alygizakis et al., 2019; López-Ruiz et al., 2019; Stastny et al., 2019) and MetFrag, which is a popular program that performs faster than similar software (Ruttkies et al., 2019a; Ruttkies et al., 2019b; Longnecker & Kujawinski, 2017; Ruttkies et al., 2016). These *in silico* fragmentation methods show some disadvantages considering fragmentation processes may vary depending on small changes in the molecular structure so rules generated from the fragmentation of a known molecule might not be applied to molecules with similar structures. Moreover, general

fragmentation rules do not explain some of the *m/z* fragments that may be present in experimental MS spectra (Nguyen et al., 2018) and some of these software are computationally expensive and therefore, limited to only small molecules (Ruttkies et al., 2016).

When organic molecules produce complex MS spectra, above mentioned software may not be able to find confident fragmentation patterns. While the *in silico* fragmentation methods above mentioned are based on a series of known chemical rules, other *in silico* fragmentation software based on the application of machine learning, have been proposed as an alternative. These machine learning methods learn the fragmentation process of molecules directly from a list of experimental spectra leading to more accurate results. Among machine learning-based *in silico* fragmentation, *competitive fragmentation modelling for metabolite identification* (CFM-ID) has been developed to simulate LC-MS/MS-ESI (both positive and negative modes) (Allen et al., 2015) as well as simpler MS techniques such as GC-EI-MS (Allen et al., 2016). This code also allows calculating the complete fragmentation patterns of a given molecule, including both immediate descendants and fragments formed from these descendants in secondary reactions that take place in the quadrupole (**Figure I7**). It has been reported that CFM-ID significantly outperforms existing state-of-the-art computational methods and has been validated and tested on the NIST database, constituting a promising alternative when no reference standard is available for measurement (Allen et al., 2016; Allen et al., 2015).

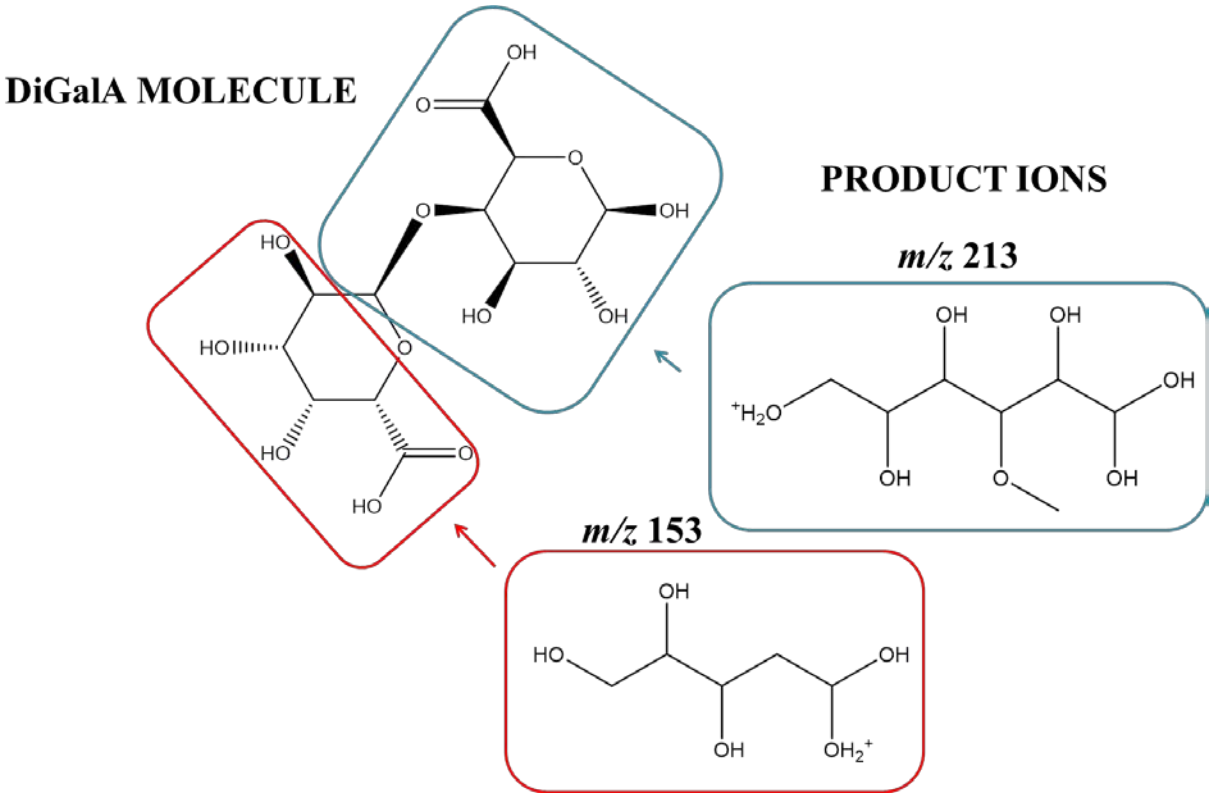


Figure I7. Example of GC-EI-MS fragments formed from digalacturonic acid (**DiGalA**) calculated using CFM-ID source code.

Some applications of *in silico* fragmentation involve generating *in silico* MS^n fragmentation pathways of small heparin oligosaccharides digest samples (DP 4) (Saad & Leary, 2005), clustering of intracellular metabolites (lyso-sulfolipids) extracted from marine diatom *Thalassiosira pseudonana* (Longnecker & Kujawinski, 2017), identification of spill components and molecular formula calculation (Alygizakis et al., 2019), fragment identification and fragmentation mechanisms of fungicides in vegetables (López-Ruiz et al., 2019) and data evaluation from mycotoxin analysis of pig colostrum and serum (Stastny et al., 2019). Few applications have been reported in the field of food science although *in silico* fragmentation models could be of great interest to elucidate structural patterns of carbohydrates and novel oligosaccharide mixtures that may have not been previously described or included in standard libraries, such as POS.

Justification and aim of the research

4. JUSTIFICATION AND AIM OF THE RESEARCH

There is a growing interest in finding alternative ways for the valorization of industrial by-products generated from vegetable processing, which could be used a source of biologically active carbohydrates like pectin that may be incorporated in functional food formulation. Among vegetable processing industries, globe artichoke (*Cynara scolymus*) has a great market value in the Mediterranean area and its manufacture generates high amounts of by-products, up to 60% of the harvested artichoke. In this sense, it has been reported that these by-products are a promising source of biologically-active ingredients. Pectins used in the industry are usually extracted from citrus peel, apple and sugar beet pulp by-products. However, efforts have been made on isolating pectin from alternative sources that may show different structural features leading to different bioactivity and technological properties. Taking into account the state of the art above indicated, the main objectives of this PhD Thesis (developed in the framework of the project "AGL2014-53445-R") were **to obtain and characterise pectin and POS from artichoke industrial by-products using enzymatic processes as an environmentally friendly technique** and **to assess the use of these products as functional ingredients by evaluating their behaviour in a dairy matrix and their potential anti-inflammatory activity**. In order to achieve these general objectives, the following partial objectives have been established:

Artichoke pectin extraction and characterization following enzymatic and ultrasound assisted methods

a) To optimise pectin extraction from artichoke by-products using a cellulase commercial preparation.
b) To compare the enzymatic method with acid and ultrasound-assisted extraction in combination with cellulase.
c) To structurally characterize artichoke pectin extracted by different methods (molecular weight distribution, monomeric composition, degree of methyl-esterification).

d) To develop an advanced data analysis strategy based on the combination of machine learning models and *in silico* fragmentation to determine structural patterns in the mass spectra of pectin extracted by different methods.

Enzymatic obtainment of POS and characterization

e) To study pectin depolymerization using several commercial enzyme preparations.
f) To compare POS obtained from different sources: high methyl-esterified pectin (citrus, apple) and low methyl-esterified pectin (artichoke).
g) To elucidate structural features and GC-EI-MS fragmentation patterns of POS present in reaction mixtures following the computational data analysis strategy developed in objective d).

Elaboration of functional yogurts with artichoke pectin/POS

h) To study the behavior of POS and other carbohydrates during the elaboration of functional yogurts containing citrus and artichoke pectin and POS.
i) To elucidate POS structures released during fermentation using computational methods.

Anti-inflammatory study of artichoke pectin and modified pectin

j) To evaluate the anti-inflammatory activity of artichoke pectin and modified pectin in dextran sulfate sodium model of mice colitis and to study cytokine expression profiles.

Work plan and structure of the thesis

5. WORK PLAN AND STRUCTURE OF THE THESIS

This Thesis comprises a multidisciplinary study and aims to make a contribution to the field of pectin and of its pectic oligosaccharides (POS) derivatives by describing pectin extraction from a novel source (artichoke bracts, stems and leaves) and performing a comprehensive characterization of its hydrolysis products to better understand their bioactivities, thus allowing to increase its functionality and expanding its potential use in the industry. To achieve the goals outlined, the work plan followed during this PhD Thesis is schematically represented in **Figure W1**.

The core of this PhD Thesis has been organized in six different chapters. The first chapter (**Articles I** and **II**) describes the optimization of pectin extraction from industrial artichoke by-products using a cellulase preparation, which was then compared with conventional extraction using acids and alternative extraction methods based on the combination of cellulase enzyme and ultrasounds. Moreover, a comprehensive structural characterization of the different pectins obtained has been carried out. In the second chapter (**Articles III** and **IV**), POS are obtained from different pectin sources (citrus, apple and artichoke) by enzymatic depolymerization using several commercial preparations with different activities. To elucidate the complex structural patterns of these POS mixtures from their mass spectra, different computational models have been developed. Lastly, the third chapter (**Articles V** and **VI**) shows two practical applications of the products obtained in the previous chapters: elaboration of functional yogurts containing citrus and artichoke pectin and POS, determining the evolution of POS during fermentation, and *in vivo* assessment of the anti-inflammatory activity of artichoke pectin and modified pectin in a murine colitis model, corroborated by establishing expression profiles of biochemical markers. Each of these three chapters correspond to two scientific papers (published or submitted) generated from this PhD Thesis and are presented in the format of the publications: abstract, introduction, materials and methods, results and discussion and conclusions. It should be noted that conclusions of each study are presented in a short section that may stand alone or form a subsection of the results and discussion section. References are place at the end of dissertation. In addition, supplementary material to each article can be found in *Annex I (I.I – I.VI)*.

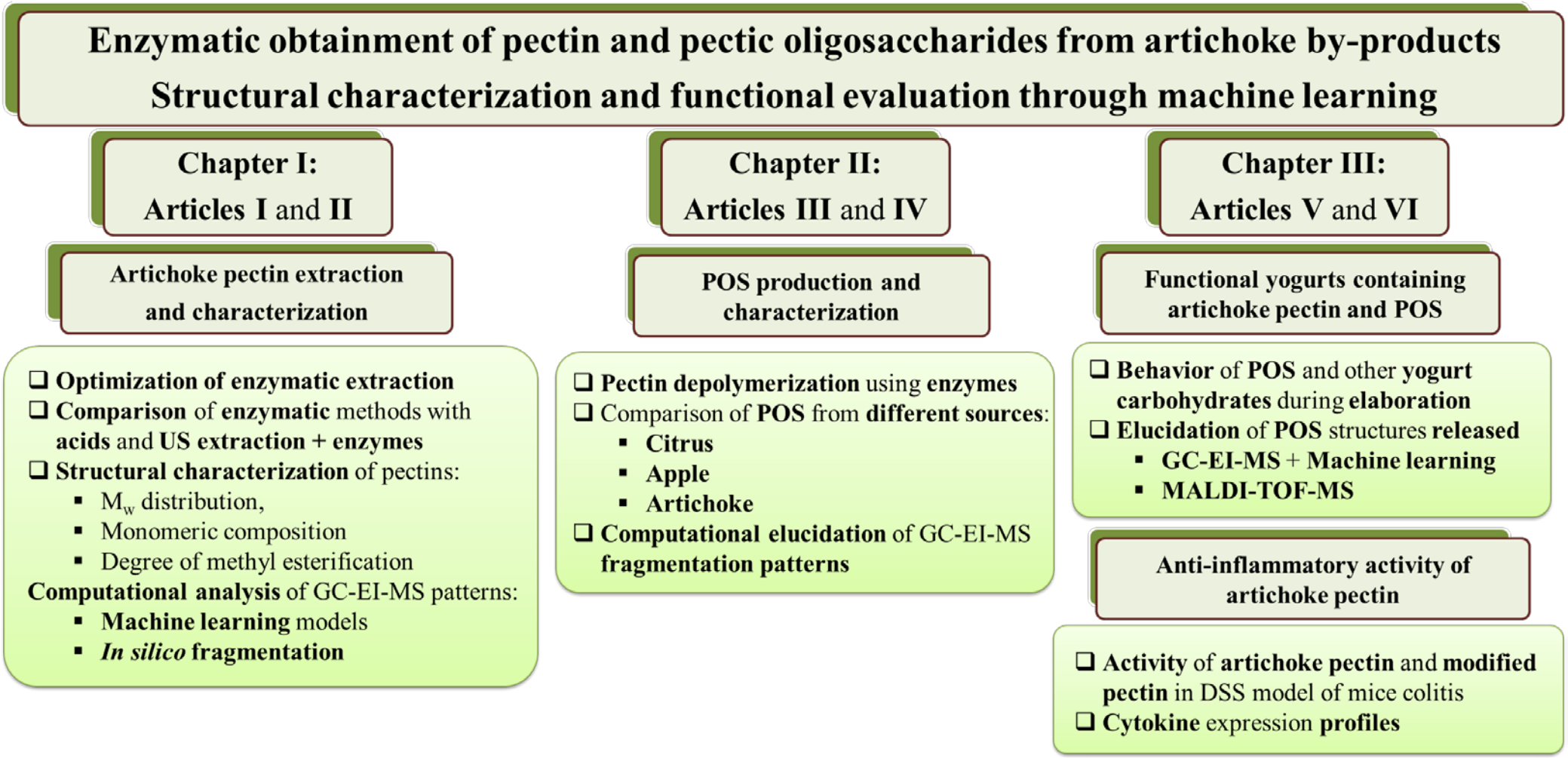


Figure W1. Schematic representation of the work plan of this PhD Thesis.

Results and discussion of the research

6. RESULTS AND DISCUSSION OF THE RESEARCH

Chapter I

Artichoke pectin extraction and characterization

Article I – Enzymatic extraction of pectin from artichoke (*Cynara scolymus L.*) by- products using Celluclast®1.5L
Article II – Ultrasound-assisted extraction of pectin from artichoke by-products. An artificial neural network approach to pectin characterization

Chapter II

POS production and characterization

Article III – GC–MS characterization of novel artichoke (*Cynara scolymus*) pectic- oligosaccharides mixtures by the application of machine learning algorithms and competitive fragmentation modelling
Article IV – Enzymatic Production and Characterization of Pectic Oligosaccharides Derived from Citrus and Apple Pectins: A GC-MS Study Using Random Forests and Association Rule Learning

Chapter III

Practical applications of artichoke pectin and POS

Article V – GC-EI-MS and MALDI-TOF-MS characterization of the carbohydrate fraction of functional yogurts containing pectin and pectic oligosaccharides. Applications of structure-retention time relationships and convolutional networks
Article VI – Intestinal anti-inflammatory effects of artichoke pectin and modified pectin fractions in dextran sulfate sodium model of mice colitis. Artificial neural network modelling of inflammatory markers

Chapter I:

Artichoke pectin extraction and characterization

Article I

6.1. Chapter I: Artichoke pectin extraction and characterization

6.1.1. Article I: Enzymatic extraction of pectin from artichoke (Cynara scolymus L.) by- products using Celluclast®1.5L

Carbohydrate Polymers 190 (2018) 43–49

Contents lists available at ScienceDirect

Carbohydrate Polymers

journal homepage: www.elsevier.com/locate/carbpol

Enzymatic extraction of pectin from artichoke (*Cynara scolymus L.*) by- products using Celluclast®1.5L

Carlos Sabater, Nieves Corzo, Agustín Olano, Antonia Montilla

Instituto de Investigación en Ciencias de la Alimentación CIAL, (CSIC-UAM) CEI (UAM + CSIC), C/Nicolás Cabrera, 9, E-28049 Madrid, Spain

Link: https://www.sciencedirect.com/science/article/pii/S0144861718302133

A B S T R A C T

The aim of this study was to optimise pectin extraction from artichoke by-products with Celluclast®1.5L using an experimental design analysed by response-surface methodology (RSM). The variables optimized were artichoke by-product powder concentration (2–7%, X_1), enzyme dose (2.2–13.3 U g^{-1}, X_2) and extraction time (6–24 h, X_3). The variables studied were galacturonic acid (GalA) (R^2 93.9) and pectic neutral sugars (R^2 92.8) content and pectin yield (R^2 88.6). In the optimum extraction conditions (X_1 = 6.5%; X_2 = 10.1 U g^{-1}; X_3 = 27.2 h), pectin yield was 176 mg g^{-1} dry matter. Considering 27.2 h of treatment as the +α value

given by the design, the extraction time was increased up to 48 h obtaining a yield of 221 mg g^{-1} dry matter. The enzymatic method optimized allows obtaining artichoke pectin with good yield, high GalA (720 mg g^{-1} dry matter) and arabinose (127.6mgg^{-1} dry matter) contents and degree of methyl-esterification of 19.5%.

Keywords: Pectin extraction, Artichoke, *Cynara scolymus*, By-products, Celluclast, RSM

INTRODUCTION

Pectins are a type of soluble fibre commonly present in vegetal cell wall and are one of the most structurally complex families of polysaccharides in nature. It is mainly composed of a linear chain of α-1,4-Dgalacturonic acid (GalA) called homogalacturonan (HG) which comprises approximately 70% of pectin. This polymer is partially methyl-esterified at the C-6 carboxyl and may be O-acetylated at O-2 or O-3. HG is the backbone that presents ramified chains of rhamnogalacturonan type I (RG-I), a polymer made up of alternate sequences of GalA and α-(1, 2) linked L-rhamnosyl residues, and rhamnogalacturonan II (RG-II), a most complex chain consisting of 12 different types of sugars in over 20 different linkages (Gullón et al., 2013; Mohnen, 2008; Garna et al., 2006). Pectins are widely used as valuable functional food ingredients due to their technological and biological properties. The majority of pectins used are extracted from citrus peel, apple and sugar beet by-products (Santos et al., 2013). Functional properties, which depend on structure of pectins, are influenced by sources and conditions of extraction, location and other environmental factors (Fissore et al., 2014; Gullón et al., 2013; Canteri et al., 2012). In the last years non-traditional sources of pectins have also been studied such as fresh peas pomace, sunflower heads, bark of mango tree, red fruit pulps and sisal wastes (Santos et al., 2013). In the same way, other new source of pectin could be artichoke by-products. Artichoke (*Cynara scolymus* L.) is an edible vegetable widely consumed in the Mediterranean diet, harvesting in this area the 85% of world production. The industrial processing of artichoke produces a high amount of wastes (external bracts, leaves and stems), which is

close to 60% of the harvested artichoke (Llorach et al., 2002). Although artichoke residues are generally used for animal feed or as manure (Machado et al., 2015), they can be used to extract bioactive compounds such as fructo-oligosaccharides and inulin (Terkmane et al., 2016; Schütz et al., 2006; López-Molina et al., 2005); flavonoids and phenolic compounds (Sałata & Gruszecki, 2010; Llorach et al., 2002) and inositols (Ruiz-Aceituno et al., 2016) as well as to isolate fractions enriched in soluble fibre (Fissore et al., 2014). However, there are scarce studies about extraction and characterization of pectin from artichoke by-products (Fissore et al., 2014; Femenia et al., 1998; Fishman et al., 1991).

The procedures for the extraction of commercial pectins are usually optimized to enhance the content of HG regions which generate the useful functional gelling properties of pectin (Maxwell et al., 2012). With this finality, the chemical extraction of pectins, employing strong acids such as hydrochloric, nitric, and sulphuric acids and high temperature (70–100 °C), has been the most widely used method for the industry (Min et al., 2011; Yapo, 2009) and although is an efficient and economical procedure, may be environmentally hazardous and could modify the structure of pectin.

Extraction of pectins can also be performed through the use of enzymes and specifically cellulases and xylanases which give rise to pectins without structural modifications. The type of enzyme could determine the yield of extraction and the composition of pectin (Wikiera et al., 2015a; Dominiak et al., 2014). Among commercial enzymatic preparations with cellulase activity, Celluclast®1.5L (Celluclast) led to complete pectin extraction with hardly any modification in its structure; thus, this enzyme has been used to extract pectin from apple pomace due to its xylano-, cellulo-, and mannolytic activities (Wikiera et al., 2015a). In addition, pectin extracted with Celluclast contained more GalA of higher degree of methyl- and acetyl-esterification compared to pectin extracted with acids (Wikiera et al., 2015b). Therefore, the aim of this study was to optimize pectin extraction from by-products (external bracts, leaves and stems) of artichoke (*Cynara scolymus*) industrialization using the commercial enzymatic preparation Celluclast. For this purpose, a three variable central composite orthogonal design analysed by response-surface methodology (RSM) was developed.

MATERIALS AND METHODS

Standards and samples

Analytical reference substances such as sucrose, D-arabinose, L-rhamnose, D-galactose, D-mannose, D-glucose, D-fructose, galacturonic acid (GalA), and β-phenyl glucoside were purchased from Sigma Aldrich (Steinheim, Germany). Eight commercial multienzymatic preparations were studied. Celluclast®1.5L (cellulase from *Trichoderma reesei*), Pectinex Ultra SP-L® (pectinase from *Aspergillus aculeatus*), Viscozyme® L (endo-1,3(4)-β-glucanase from *A. aculeatus*), Viscozyme® Barley HT (amylase and cellulase from *Aspergillus* sp.) and Carezyme® 4500 L (cellulase from *Aspergillus* sp.) were generous gifts from Novozymes (Bagsvaerd, Denmark). Pectinase and cellulase from *A. niger* and cellulase from *T. viride* were acquired from Sigma. Cellulase and pectinase activities were measured by quantifying the amount of reducing sugars released, using the dinitrosalicylic acid (DNS) method (Miller, 1959). To determine cellulase activity, cellulose solutions (10 mg mL^{-1}, Carboxymethyl cellulose, Sigma) dissolved in sodium acetate buffer (0.05 M, pH 5) with 5 μL mL^{-1} or 1 mg mL^{-1} of enzyme at 50 °C for 5 min were prepared and glucose was employed as standard. One unit of cellulase activity was defined as the amount of the enzyme, which released 1 μmol of glucose per min. Pectinase activity was measured using polygalacturonic acid solutions (10 mg mL^{-1}, Sigma) incubated with 5 μLmL^{-1} or 1 mg mL^{-1} of enzyme at 50 °C for 2 min. GalA was employed as a standard. One unit of pectinase activity was defined as the amount of the enzyme, which released 1 μmol of GalA per min. Industrial by-product of artichoke, consisting in blanched external bracts, leaves and stems, were kindly supplied by Riberebro® (La Rioja, Spain). Prior to characterization, this product was freeze-dried, homogenized using a Wiley Mill 934 equipment (Thomas Scientific, USA) and sieving through a 500 μm sieve (Femenia et al., 1998).

Pectin extraction from artichoke by-products

Enzymatic extraction of pectin from artichoke by-products (particle size ≤ 500 μm) was performed using Celluclast. The process was conducted in an orbital shaker at 50 °C, pH 5 with constant shaking (200 rpm). After hydrolysis, samples were centrifuged (1300g, 10 min, 4 °C) and supernatants were filtered through cellulose paper. The resultant precipitate was washed with distilled water (20 mL), centrifuged and filtered. The supernatants were combined and ethanol at 4 °C was added to obtain a final concentration of 70%. After one hour at 4 °C, the precipitated pectins were centrifuged (1200g, 20 min), washed twice with 70% ethanol (600 mL), centrifuged again, and freeze-dried.

Experimental design for extraction of pectins

A three level, three variable, central composite orthogonal design was applied to determine the best conditions to extract enzymatically pectin from artichoke by-products. The factorial design consisted of 8 factorial points, 6 axial points (two axial points on the axis of each design variable at a distance of 2 from the design centre) and 3 centre points. The independent variables used in this study were artichoke byproduct powder concentration (2–7% w/w, X_1), enzyme dose (2.2–13.3 U g^{-1}, X_2), and extraction time (6–24 h; X_3). Four dependent variables were taken into account to optimise pectin extraction by means of RSM: yield of pectin extraction (mg of extracted pectin/g dry matter, Y_1), GalA content (mg g^{-1} dry matter, Y_2), pectic neutral sugars content (arabinose, rhamnose and galactose, mg g^{-1} dry matter, Y_3) and other neutral sugars content (glucose, mannose and fructose, mg g^{-1} dry matter, Y_4).

The responses obtained from each set of experimental design were subjected to multiple nonlinear regressions to achieve the coefficients of the second order polynomial model. The quality of the fit of the polynomial model equation was expressed by the coefficient of determination R-squared (R^2) and the adjusted R-squared (R^2adj), which provides a measurement of how much of the variability in the observed response values could be explained by the experimental factors and their interactions (Myers et al., 2004). Variables that presented R^2adj values higher than 70%

were submitted to simultaneous optimization. The overall effect of the three independent factors was used to obtain a desirability function that represents the influence of the extraction conditions on the efficiency of enzymatic extraction with Celluclast and on the final product composition. In general terms, a desirability function represents the effect of the independent variables (X_i) on the dependent variables (Y_j), determining the process efficiency (Myers et al., 2004). For each extraction conditions, the desirability function assigns numbers between 0 (completely undesirable value) and 1 (completely desirable or ideal response). The method allows optimize extraction conditions that provide the "most desirable" response values. Additional extractions of artichoke pectin were carried out at following conditions: artichoke by-product powder 6.5% (w/w), enzyme dose 10.1 U g^{-1} and extraction time 36 and 48 h.

Characterization of samples

All analyses were carried out in duplicate.

Artichoke by-products and extracted pectin

Dry matter was determined gravimetrically drying the samples until constant weight according to the AOAC method 950.01 (Association of Official Analytical Chemists, 1990a).

pH was measured in solutions of 10% using a Mettler Toledo Five Easy Plus pH Meter.

Water activity (a_w) measurement was carried out in an AW Sprint TH-500 instrument (Novasina, Pfäffikon, Switzerland). Saturated aqueous solutions of LiCl, $MgCl_2$, $Mg(NO_3)_2$, NaCl, $BaCl_2$ and $K_2Cr_2O_7$ were used to calibrate the sensor unit.

Protein was determined following the Kjeldhal method for total nitrogen (TN) using a conversion factor (TN × 6.25) to express results as protein (Association of Official Analytical Chemists, 1990b).

Artichoke by-products and extracted pectin

Fat content was measured following AOAC 920.39 (Association of Official Analytical Chemists (AOAC), 1990c) methodology performing Soxhlet extraction using a Gerhardt soxtherm extractor (Germany).

Water soluble (WSF) and insoluble (WIF) fractions were obtained by immersing 1.5g of powdered samples in 40mL boiling water for 30min. The mixture was cooled and filtered through a sintered glass filter (No 2). The residue was re-extracted twice with boiling water and then washed with 40 mL of water. Supernatant was collected and then residue was washed with 40 mL of acetone 70% and 40 mL of ethanol. Water insoluble fraction (WIF) was dried and stored for further analysis. Supernatant (water soluble fraction, WSF) was partially dried at 38–40 °C in a rotary evaporator (Büchi Labortechnik AG, Flawil, Switzerland), freeze-dried, and redissolved in water at 10 mg mL^{-1} for chromatographic analysis.

Alcohol soluble (ASF) and insoluble (AIF) fractions were obtained by immersing 1.5 g of powdered samples in 40 mL boiling ethanol, final concentration 80% (v/v), for 30 min. The mixture was cooled and filtered through a sintered glass filter (No 2). The residue was re-extracted twice at the same conditions. Supernatants were collected and then residue was washed first with 40 mL of 70% acetone and then with 40 mL of ethanol. Alcohol insoluble fraction (AIF) was dried and stored for further analysis. Supernatant (alcohol soluble fraction, ASF) was partially dried at 38–40 °C in a rotary evaporator, freeze-dried and redissolved in water at 10 mg mL^{-1} for chromatographic analysis.

Analytical techniques

ICP-MS

The ion composition of artichoke by-product powder and extracted pectin was determined using an ICP-MS ELAN 6000 Perkin-Elmer Sciex instrument at the Servicio Interdepartamental de Investigación (SIdI-UAM) of Madrid. Quantitative analysis of the

elements of interest using the external calibration method and internal standard to correct instrumental drift was carried out.

FT-IR

In order to determine the degree of methyl-esterification, freeze-dried samples of artichoke by-products and extracted pectin were analysed by FT-IR analysis. KBr discs were prepared mixing the pectin samples with KBr (1:100) and pressed. FT-IR spectra were performed in a Bruker IFS66 v equipment (Bruker, US). Data were collected in absorbance mode using a frequency range of 4000–400 cm^{-1}, and resolution of 4 cm^{-1} (mid infrared region) with 250 co-added scans. Degree of methyl-esterification was determined as the average of the ratio of the peak area at 1747 cm^{-1} (COO-R) over the sum of the peak areas of 1747 cm^{-1} (COO-R) and 1632 cm^{-1} (COOe) as described earlier (Singthong et al., 2004).

Gas chromatography —FID

Sample preparation. To determine the monomeric composition, artichoke by-product powder was hydrolysed using sulfuric acid ($H_2SO_{4)}$ following the methods of Seaman (Yapo & Koffi, 2008) and trifluoroacetic acid (TFA) (Garna et al., 2006). Briefly, in the first method 35 mg of sample were dispersed in 300 μL of 72% H_2SO_4 at room temperature for 3 h, diluted with 1 M H_2SO_4 and heated at 100 °C for 2.5 h. Supernatant was dried at 38–40 °C in a rotary evaporator for chromatographic analysis. For hydrolysis with TFA 30 mg of sample were hydrolysed with 1.5 mL of TFA 2 M at 110 °C under inert conditions for 4 h (Garna et al., 2006).

In order to determine the monomeric composition of extracted artichoke pectin, powder samples were hydrolysed with TFA following the method mentioned above. Similarly, standards of $_D$-arabinose, $_L$- rhamnose, $_D$-fructose, $_D$-galactose, $_D$-mannose, $_D$-glucose and GalA at 20 mg mL^{-1} were hydrolysed to correct the obtained results.

Besides, enzymatic hydrolysis of artichoke pectin, obtained in the optimal conditions, has been carried out to determine GalA content. Samples (2% w/v) were

dissolved in 0.05 M sodium acetate buffer (pH 5.0) and hydrolysed with 90 U mL^{-1} of Viscozyme® L preparation (Garna et al., 2006). Then, samples were incubated in an orbital shaker at 50 °C and 750 rpm (0.94 *g*) for 24 h and immediately immersed in boiling water for 5 min to inactivate the enzyme.

Chromatographic analyses. Monomeric composition of artichoke by-product powder and extracted pectins was determined by GC-FID as trimethyl silylated oximes (TMSO) formed following the method of Brobst and Lott (1966). First, samples containing between 2.5 at 5 mg of sugars were added 0.4 mL of internal standard (I.S.) solution (0.5 mg mL^{-1} phenyl-β-glucoside). Afterwards, the mixtures were dried at 38–40 °C in a rotary evaporator. Sugar oximes were formed by adding 250 μL of hydroxylamine chloride (2.5%) in pyridine and heating the mixture at 70 °C for 30 min and then silylated with hexamethyldisilazane (250 μL) and TFA (25 μL) and kept at 50 °C for 30 min. Reaction mixtures were centrifuged at 6708*g* for 2 min at room temperature. Supernatants were injected or stored at 4 °C prior to analysis.

GC-FID analysis of TMSO derivatives was performed on an Agilent Technologies 7890A gas chromatograph (Wilmington, DE, USA) using a fused silica capillary column DB-5HT, bonded, crosslinked phase (5% phenyl-methylpolysiloxane; 15 m × 0.32 mm i.d., 0.10 μm film thickness) (J&W Scientific, Folson, California, USA). The oven temperature was initially 150 °C increased at a rate of 1 °C min^{-1} to 165 °C, then increased at a rate of 10 °C min^{-1}–200 °C, then increased at a rate of 50 °C min^{-1} to 380 °C and held for 2 min. The injector and detector temperatures were set at 280 and 385 °C, respectively. Injections were carried out in split mode (1:5) using nitrogen at 1 mL min^{-1} as carrier gas. To study the response factors relative to the internal standard, solutions containing arabinose, rhamnose, galactose, mannose, glucose, GalA, raffinose and stachyose were prepared over the expected concentration range in samples. The identities of carbohydrates were confirmed by comparison with relative retention times of standard samples. Data acquisition and integration were performed using Agilent ChemStation Rev. B.03.01 software. All analyses were carried out in duplicate and data were expressed as mean ± standard deviation (SD).

High performance size-exclusion chromatography with evaporative light scattering detector (HPSEC-ELSD)

The molecular weight (M_w) distribution of carbohydrates present in the reaction mixtures obtained by Celluclast activity over artichoke by-product was determined by HPSEC-ELSD. Samples (0.65 mg mL^{-1}) were dissolved in water, filtered using a 0.45 μm syringe filter (Symta, Spain) and analysed in an Agilent Technologies 1220 Infinity LC System (Böblingen, Germany) equipped with an evaporative light scattering detector (1260 Infinity ELSD). The nebulization temperature was 75 °C. Synthetic air at 85 °C was used for the evaporation at a flow-rate of 1.2 mL min^{-1}. The separation of carbohydrates was carried out on a TSK-GEL G5000PW_{XL} column (300 mm × 7.8 mm, 10 μm particle size) and TSK-GEL G2500PW_{XL} column (300 mm × 7.8 mm, 6 μm particle size) in combination with a TSK-GEL PW_{XL} guard column (40 mm × 6 mm, 12 μm particle size) (Tosoh Bioscience, Montgomeryville, PA, USA) using 0.01 M NH_4Ac, as mobile phase and elution in isocratic mode at a flow rate of 0.5 mL min^{-1} for 50 min. The column was thermostatized at 25 °C and the injection volume was 50 μL (~32 μg of total carbohydrates). Data acquisition and processing were performed employing Agilent ChemStation software.

M_w of carbohydrates was calculated by the external calibration method using solutions of commercial pullulan standards (M_w 0.342–788 kDa) (Fluka Analytical) in the range 10–2250 mg L^{-1}. All analyses were performed in duplicate, obtaining relative standard deviation (RSD) values below 10% in all cases.

Statistical analyses

Experimental design, RSM and statistical analysis were performed using the software STATGRAPHICS Centurion XVI.I. (Statistical Graphics Corporation, Rockville, MD, USA). Other data not obtained in the experimental design were submitted to an analysis of variance ($p < 0.05$).

RESULTS AND DISCUSSION

Characterization of artichoke by-product

The values of pH, dry matter, a_w, protein, fat, degree of methyl-esterification, minerals and different carbohydrate fractions (AIF, ASF, WIF and WSF) found in industrial artichoke by-product powder are listed in Table R1.1. Protein, fat and mineral contents were similar to reported by USDA (2017) for artichokes cooked, boiled and drained. Water and alcohol insoluble fractions (WIF and AIF), represented 926 and 928 mg g^{-1} dry matter. These values were higher than that found by Femenia et al. (1998) (64% for external bracts), however this difference could be explained by the possible leaching occurred during artichoke by-product blanching, as was indicated by Gamboa-Santos et al., (2012) that reported a loss of weight up to 31% of soluble solids in the blanching water of carrot samples subjected to treatment at 95 °C during 5 min. WSF and ASF represented 74 and 72 mg g^{-1} dry matter of the sample, respectively. Carbohydrate composition of WSF and ASF was determined by GC-FID, and as it can be seen in Table R1.2 the most abundant low M_w carbohydrates were fructose, glucose, sucrose, kestose and nystose, most of them probably released by the spontaneous hydrolysis of inulin, occurred during extraction of WIF and AIF. Besides, artichoke by-product powder was hydrolysed with acids to know monomeric composition; the results are depicted in Table R1.2. As observed hydrolysates using H_2SO_4 (Seaman hydrolysis) contained arabinose, rhamnose, galactose and galacturonic acid (GalA) as pectic monosaccharides; and fructose, mannose and glucose indicating the presence of other polysaccharides as fructans, mannans, cellulose and hemicellulose in by-products from artichoke. The same type of monosaccharides was found in TFA hydrolysates, but in higher amounts than in H_2SO_4 ones, except for fructose and glucose.

Table R1.1. Characterization of powder artichoke by-product and extracted pectin.

	Artichoke by-product powder Mean ± SD	Extracted artichoke pectin Mean ± SD
pH	4.9	6.0
Dry matter (%)	91.7 ± 2.5	90.5 ± 0.3
a_w	0.20 ± 0.01	0.26 ± 0.00
Protein (mg g^{-1} dry matter)	140.5 ± 15.8	20.7 ± 1.0
Fat (mg g^{-1} dry matter)	2.4 ± 0.2	–
Alcohol insoluble fraction (AIF) (mg g^{-1} dry matter)	928.2 ± 5.9	–
Alcohol soluble fraction (ASF) (mg g^{-1} dry matter)	71.8 ± 5.9	–
Water insoluble fraction (WIF) (mg g^{-1} dry matter)	926.5 ± 1.6	–
Water soluble fraction (WSF) (mg g^{-1} dry matter)	73.5 ± 1.6	–
K (mg g^{-1} dry matter)	16.9	10.6
Na (mg g^{-1} dry matter)	3.5	9.6
Ca (mg g^{-1} dry matter)	4.7	7.3
Mg (mg g^{-1} dry matter)	2.4	2.9
B (mg g^{-1} dry matter)	0.02	0.04
Degree of methyl-esterification (%)	9.1 ± 0.5	19.5 ± 0.0

GalA levels in TFA hydrolysates were about 34 mg g^{-1} dry matter, fifteen times higher than using H_2SO_4. Since GalA is the main constituent of pectin, the high content found in TFA hydrolysates may indicate that the artichoke by-product is a good source to obtain pectin. In addition, taking into account the results of both hydrolysis, acid hydrolysis with TFA was selected to determine the monomeric composition of pectin extracted in the different assays of the experimental design with the aim of optimise the conditions of enzymatic extraction of artichoke pectin.

Optimization of enzymatic extraction of pectin from artichoke by-product

In order to select the best enzyme to extract pectin from artichoke by-product, different commercial preparations were tested (Table S1.1). According to the results obtained, Celluclast presented high values of cellulase activity although pectinase activity was not detected, then it is expected that extracted pectin will not be degraded. On the other hand, Viscozyme®L presented the highest pectinase activity. Therefore, this enzyme was selected for enzymatic hydrolysis of pectin to determine GalA content of most accurate form. Once Celluclast preparation was selected for pectin extraction, the experimental design was developed to optimize the extraction conditions using independent variables: artichoke by-product powder concentration (X_1), enzyme dose (X_2) and extraction time (X_3). Table R1.3 shows the results obtained during the enzymatic extraction of pectin from artichoke by-product for the different experimental variables studied. The yields of extracted pectin (Y_1) in the 17 assays of experimental design were in range of 98.1–208.3 mg g^{-1} dry matter. The highest pectin yield was obtained in a reaction with 4.5% of artichoke by-product powder, 7.8 U g^{-1} of enzyme during 27.2 h. This value was close to the theoretical optimum 225.7 mg g^{-1} dry matter obtained, according to the RSM analysis, after 27.2 h using 7.9% of substrate and 10.7 U g^{-1} of enzyme.

Following with the optimization of pectin extraction design using Celluclast, other dependent variables, as GalA content (Y_2), pectic neutral sugars (Y_3, arabinose, rhamnose and galactose) and other neutral sugars (Y_4, fructose, glucose and mannose) were determined (Table R1.3). Hydrolysis of extracted artichoke pectin using TFA was

performed to obtain individual values of each monosaccharide (see Table S1.2). The main carbohydrate quantified was GalA (Y_2) being in the range of 30.5–58.1 mg g^{-1} dry matter reaching the maximum value at the same conditions of substract concentration (4.5%) and enzyme dose (7.8 U g^{-1}) which was found the highest yield but at lower time of extraction (15 h). The content of neutral sugars characteristic of pectin (Y_3) was between 12.2–37.2 mg g^{-1} dry matter; for these compounds maximum content was achieved at the highest extraction time and low artichoke by-product powder concentrations (4.5%). During the obtainment of pectin using Celluclast, other polysaccharides as cellulose, hemicellulose, mannan or inulin were extracted being the monosaccharides released from these polysaccharides glucose, mannose and fructose; the amount of these compounds (Y_4) in the different assays was between 7.3 and 21.8 mg g^{-1} dry matter. Unlike the other parameters, the value of this variable must be minimized.

The optimization of the experimental conditions for pectin extraction and the results obtained of the most representative dependent variables under study (yield, GalA, pectic neutral sugars and other neutral sugars contents) were related by means of RSM. The regression equations describing the changes in these studied variables are shown in Table R1.4. As it can be observed, the variables that presented higher regression values were content of GalA, pectic neutral sugars, and pectin yield (R^2 = 93.9, 92.8 and 88.6%; and R^2 = 86.0, 83.6 and 73.8%, respectively). Moreover, these variables are the most important in a pectin extraction process; the extraction was aimed at maximizing pectin yield and to get a product with the highest content of pectic sugars (GalA, arabinose, rhamnose and galactose) that reflect a high purity of pectin. So, these variables were selected to optimize the enzymatic extraction process using desirability function. The corresponding three-dimensional representation of the desirability function obtained is shown in Figure R1.1. Maximum predicted desirability (0.96) corresponds to an artichoke by-product powder concentration of 6.5%, an enzyme dose of 10.1 U g^{-1} and an extraction time of 27.2 h corresponding to a pectin yield of 175.6 mg g^{-1} dry matter. Among the tested conditions, the highest observed values of desirability function were 0.86, 0.80 and 0.79 (corresponding to assays 9, 1 and 12) and 0.66 ± 0.04 to the central points (assays 3, 5 and 6).

Table R1.2. Carbohydrate composition found in water (WSF) and alcohol soluble fractions (ASF) from artichoke by-product powder and hydrolysates of artichoke by-product powder using H_2SO_4 (Seaman hydrolysis) and TFA (2 M).

Carbohydrates (mg g^{-1} dry matter)				
	Water soluble fraction (WSF)	Alcohol soluble fraction (ASF)	Hydrolysates	
			H_2SO_4	TFA
Arabinose	–	–	14.2 ± 4.4	72.0 ± 18.5
Rhamnose	–	–	1.1 ± 0.0	10.9 ± 2.2
Fructose	21.8 ± 0.0	18.5 ± 1.1	4.4 ± 4.4	3.3 ± 2.2
Galactose	1.1 ± 0.0	–	2.2 ± 1.1	17.4 ± 4.4
Mannose	6.5 ± 1.1	4.4 ± 1.1	1.1 ± 0.0	29.4 ± 6.5
Glucose	15.3 ± 1.1	14.2 ± 1.1	19.6 ± 6.5	15.3 ± 4.4
Galacturonic acid	–	–	2.2 ± 0.0	33.8 ± 7.6
Sucrose	6.5 ± 1.1	7.6 ± 0.0	–	–
Kestose	6.5 ± 0.0	6.5 ± 0.0	–	–
Nystose	8.7 ± 1.1	12.0 ± 1.1	–	–

Taking into account that optimal extraction time (27.2 h) was the +α value given by the design and probably the maximum of pectin extraction may not have been reached, extraction time of pectin was increased at 36.0 and 48.0 h using the same optimized concentration of artichoke by-product powder (6.5%) and enzyme dose (10.1 U g^{-1}). According to the obtained results, pectin extracted after 48 h of hydrolysis contained more GalA and higher yields were obtained. Under the above mentioned conditions, the experimental yield was 22.1% (221.4 mg g^{-1} dry matter), which is well matched with the maximum predictive yield (225.7 mg g^{-1} dry matter). Therefore, the use of enzymatic preparation Celluclast is a good method to extract pectin from artichoke by-products and it could be an alternative to traditional methods which use acids to extract pectin with high yield. Dominiak et al. (2014) extracted pectin from lime peel with different cellulases obtaining yield in the range of 18–26%, while using nitric acid the pectin yields were between 13 and 26%; similarly, Wikiera et al. (2015a) obtained higher yield with Celluclast (19%) than in sulfuric acid (15%) using apple pomace as raw material. Liew et al., (2016) obtained higher pectin yield from passion fruit peels with Celluclast (9.2%) than with citric acid (7.7%).

Characterization of extracted artichoke pectin

Finally, a general characterization of extracted artichoke pectin was performed. Thus, the physico-chemical parameters such as pH, dry matter, a_w, protein and degree of methyl-esterification were determined and the results are shown in Table R1.1. As can be observed the protein content of extracted pectin was low enough (20.7 mg g^{-1} dry matter), therefore it is not necessary to use any protease during enzymatic extraction to avoid presence of high level of protein in extracted pectin. With respect to mineral content in comparison with artichoke by-product, the most abundant cation was K, although its content was lower in the extracted pectin, besides, Na content was also greater probably due to buffer used during extraction. High Ca concentrations were found, probably due to high affinity of pectin to this divalent cation. Also, it is important to point out the high content of boron 0.04 mg g^{-1} dry matter, associated to the pectin (the content of artichoke by-product powder was 0.02 mg g^{-1} dry matter). In the bibliography has been reported that rhamnogalacturonan II (RG-II) suffer reversible dimerization with borate contributing to wall strength of the plant cell (Caffall & Mohnen 2009), similar role have calcium in apple and citrus pectin, while, in these products boron was not detected (Muñoz-Almagro et al., 2017). The artichoke pectin obtained was classified as low-methoxyl pectin with a degree of methyl-esterification of 19.5%, similar to other pectins from sunflowers and lower than those of apple and citrus pectin previously characterized in our laboratory (72–76%). Other parameter studied in extracted pectin was the Mw distribution analysed by HPSEC-ELSD.

Table R1.3. Values of independent variables (artichoke by-product powder concentration X_1, enzyme dose X_2 and extraction time X_3) and dependent variables (pectin yield Y_1, galacturonic acid content Y_2, pectic neutral sugars Y_3 and other neutral sugars Y_4), and predicted and observed desirability found in each assay of pectin extraction after Celluclast treatment of artichoke by-products. Data are expressed as mg g^{-1} dry matter artichoke by-product powder.

Run	X_1, artichoke by-product powder concentration (%)	X_2, Enzyme dose (U g^{-1})	X_3, Extraction time (h)	Y_1, Pectin yield (mg g^{-1})	Y_2, GalA (mg g^{-1})	Y_3, Pectic neutral* sugars (mg g^{-1})	Y_4, Other neutral** sugars (mg g^{-1})	Predicted desirability	Observed desirability
1	7.0 (+1)	13.3 (+1)	24.0 (+1)	190.8	49.8 ± 2.4	32.6 ± 2.7	15.8 ± 1.0	0.90	0.80
2	2.0 (−1)	13.3 (+1)	6.0 (−1)	163.6	47.8 ± 0.1	21.3 ± 0.3	11.5 ± 0.7	0.34	0.27
3	4.5 (0)	7.8 (0)	15.0 (0)	176.7	52.2 ± 0.9	28.9 ± 0.5	15.0 ± 0.4	0.68	0.62
4	4.5 (0)	0.3 (−2)	15.0 (0)	106.9	30.5 ± 0.5	13.1 ± 0.1	7.3 ± 0.2	0.10	0.00
5	4.5 (0)	7.8 (0)	15.0 (0)	179.9	58.1 ± 5.1	31.2 ± 1.7	15.5 ± 1.7	0.68	0.70
6	4.5 (0)	7.8 (0)	15.0 (0)	197.4	50.6 ± 1.2	29.8 ± 0.7	16.2 ± 1.3	0.68	0.65
7	2.0 (−1)	2.2 (−1)	24.0 (+1)	133.0	45.3 ± 0.1	22.5 ± 0.4	10.0 ± 0.5	0.00	0.00
8	4.5 (0)	15.3 (+2)	15.0 (0)	187.6	51.6 ± 2.1	30.5 ± 2.1	16.4 ± 3.4	0.55	0.65
9	4.5 (0)	7.8 (0)	27.2 (+2)	208.3	57.6 ± 1.9	37.2 ± 0.7	21.8 ± 1.2	0.81	0.86
10	7.0 (+1)	2.2 (−1)	6.0 (−1)	98.1	32.2 ± 3.1	12.2 ± 0.7	8.2 ± 0.5	0.00	0.00
11	2.0 (−1)	13.3 (+1)	24.0 (+1)	167.9	52.1 ± 3.3	36.1 ± 2.8	12.6 ± 1.6	0.40	0.41
12	7.9 (+2)	7.8 (0)	15.0 (0)	185.4	48.7 ± 1.5	32.8 ± 0.1	21.5 ± 1.1	0.73	0.79
13	2.0 (−1)	2.2 (−1)	6.0 (−1)	159.2	37.4 ± 0.4	20.4 ± 0.4	9.7 ± 1.2	0.11	0.18
14	7.0 (+1)	13.3 (+1)	6.0 (−1)	129.8	44.6 ± 0.7	23.4 ± 1.1	13.4 ± 0.5	0.50	0.50
15	7.0 (+1)	2.2 (−1)	24.0 (+1)	165.8	38.6 ± 1.3	24.6 ± 1.0	13.8 ± 0.9	0.48	0.50
16	1.1 (−2)	7.8 (0)	15.0 (0)	169.0	51.6 ± 3.4	33.5 ± 0.9	11.2 ± 0.8	0.27	0.00
17	4.5 (0)	7.8 (0)	2.8 (−2)	158.1	41.9 ± 3.4	23.3 ± 1.4	12.6 ± 1.0	0.43	0.43

Y_2, Y_3 and Y_4: values found in TFA hydrolysates of extracted artichoke pectin (See Table S1.2).

* ∑ of arabinose, rhamnose and galactose. ** ∑ of fructose, glucose and mannose.

Table R1.4. Regression equations for the model fit of the different variables studied during the enzymatic extraction of pectin from artichoke by-product using Celluclast.

Variables	Fitted model equation*	R^2 (%)	R^2 adj (%)
Pectin yield (Y_1, %)	$Y_1 = 11.8809 + 1.22659X_2 + 0.0766667X_3X_1 - 0.069827X^2$	88.6	73.8
Galacturonic acid (Y_2, mg g^{-1})	$Y_2 = 0.593232 + 0.161405X_2 - 0.00790439X_2^2$	93.9	86.0
Neutral sugars from extracted pectin (Y_3, mg g^{-1})	$Y_3 = 0.647902 + 0.191896X_2 - 0.0110103X_2^2$	92.8	83.6
Other neutral sugars (glucose, mannose and fructose content, Y_4, mg g^{-1})	$Y_4 = 0.0661304 + 0.130052X_2 - 0.00695625X^2$	77.9	49.4

* (X_1) Artichoke by-product powder concentration (%), (X_2) enzyme dose (U g^{-1}), (X_3) extraction time (h).

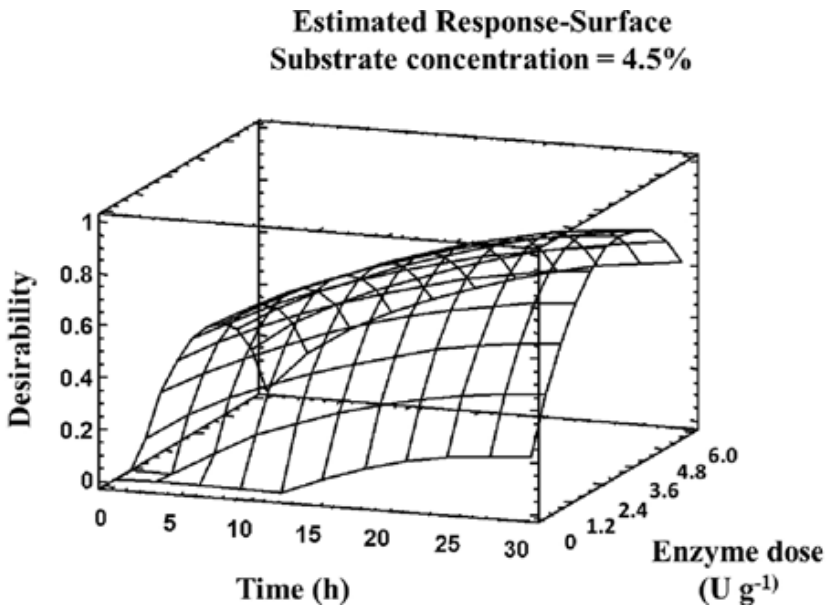


Figure R1.1. Estimated response surface for the desirability function.

The pectin extracted at optimal conditions presented the profile with three fragments of 660, 105 and 4.8 kDa, with abundances of 28.5, 36.6 and 34.9% respectively (Figure R1.2) This size distribution non monomodal is different to that of pectins extracted from citrus or apple (Muñoz-Almagro et al., 2017).

Following in artichoke pectin characterization, a determination of monomeric composition was performed in acid hydrolysates (TFA). Pectic monosaccharides such as arabinose (126.6 ± 5.8 mg g^{-1} dry matter of pectin), galactose (23.5 ± 0.4 mg g^{-1} dry matter), and rhamnose (18.8 ± 2.7 mg g^{-1} dry matter) were quantified. Besides, glucose (30.4 ± 4.4 mg g^{-1} dry matter), mannose (26.8 ± 0.8 mg g^{-1} dry matter) and fructose (4.6 ± 2.9 mg g^{-1} dry matter) were also present. Taking into account that GalA suffers a high degradation during acid hydrolysis, it was determined after enzymatic hydrolysis of pectin using Viscozyme® (Babbar et al., 2016a). GalA content was 720.0 ± 44.6 mg g^{-1} dry matter (72.0% dry matter), which is higher than 65% being the limit to consider pectin as food additive (E-440) (Morris et al., 2013). The characteristic that could be highlighted of monomeric composition of extracted artichoke pectin is the high arabinose content. This is higher

than found in pectin extracted from other sources like citrus and apple (Wikiera et al., 2015a; Gómez et al., 2013; Garna et al., 2004) and similar to pectin from sugar beet pulp (Leijdekkers et al., 2015).

The amounts of rhamnose, arabinose and galactose with respect to GalA in the purified pectin indicated possible branching along the homogalacturonan (Yuliarti et al., 2015). Considering the monomeric composition of pectin, the degree of branching (GalA/Rha), linearity pectin backbone [GalA/(Rha + Ara + Gal)] and extent of branching of RG-I [(Ara + GalA)/ Rha] were 19.2, 3.2 and 23.0, respectively. The degree and extent of branching of RG-I were higher and the linearity pectin backbone was lower than those obtained for pectins from other sources as gold kiwifruit (Yuliarti et al., 2015) or lime peel (Dominiak et al., 2014). Another influential factor that can modify the structure of pectin is the extraction method, thus Fishman et al. (1991) obtained artichoke pectin with a low content of arabinose (3.4%), rhamnose (1.4%) and galactose (1.8%) using ammonium oxalate; being lower than orange pectin extracted with the same method, this fact may indicate a loss of branches. The maintenance of the structure, and more specifically the high arabinose content, could confer to this pectin anticarcinogenic property (Wikiera et al., 2015a).

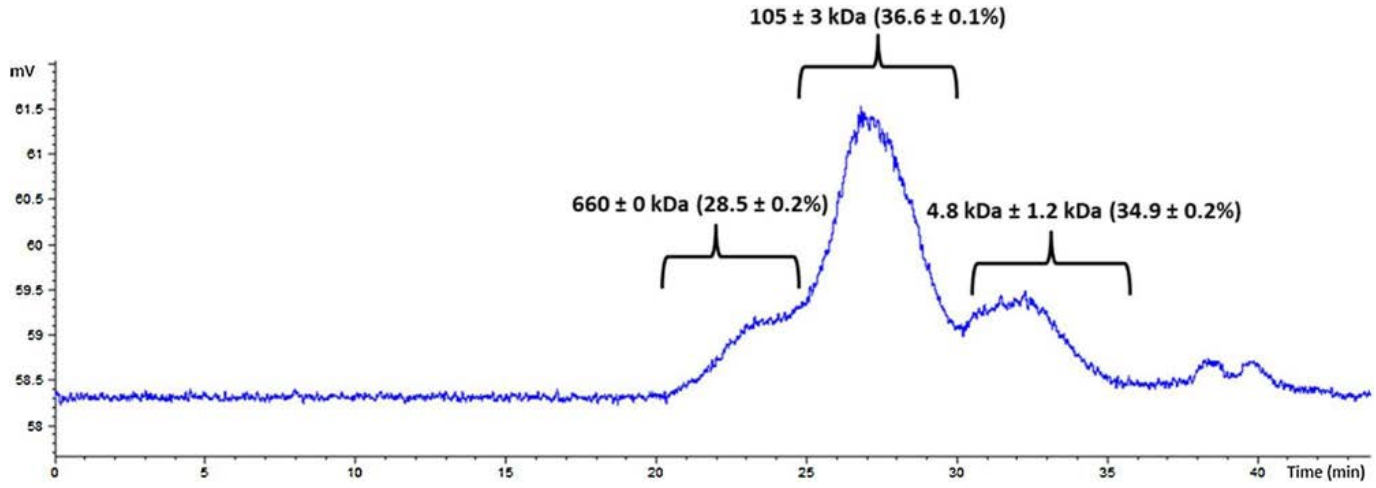


Figure R1.2. HPLC-SEC-ELSD profile of a hydrolysate of artichoke by-product obtained by enzymatic hydrolysis using Celluclast® 1.5 L under optimal conditions (artichoke by-product powder concentration 6.5%; enzyme dose 10.1 U g^{-1}; extraction time 48 h).

As it has been indicated in the introduction, few papers have been published on the production and characterization of artichoke pectin. Thus, Fishman et al. (1991) studied the monomeric composition of carbohydrate fraction extracted with ammonium oxalate from artichoke, found high amount of GalA and other typical monosaccharides of pectin. Femenia et al. (1998) reported GalA concentrations about 40% in the AIF of artichoke stems, shown to be a good source of pectic polysaccharide. Fissore et al. (2014) obtained soluble fibre extracts from artichoke by-products with a GalA content between 17 and 25% of total carbohydrates. However, no pectin separation was carried out in both cases.

CONCLUSIONS

Artichoke by-products (a mixture of stems, leaves and bracts) can be used as a good source of pectins. The extraction of pectin from these residues using commercial Celluclast preparation has been optimized allowing an efficient extraction giving rise to a yield of 221.4 mg g^{-1} dry matter. Enzymatic extraction is an environmentally-friendly process and allows to obtain high yields of pectin under optimal conditions. The pectin extracted from artichoke by-product powder was characterized finding a high GalA content (720.0 mg g^{-1}) and due to its degree of methyl-esterification (19.5%) was classified as low-methoxyled pectin. These characteristics would allow the use of this pectin as a food ingredient.

Article II

6.1.2. Article II: Ultrasound-assisted extraction of pectin from artichoke by-products. An artificial neural network approach to pectin characterization

Food Hydrocolloids 98 (2020) 105238

Contents lists available at ScienceDirect

Food Hydrocolloids

journal homepage: www.elsevier.com/locate/foodhyd

Ultrasound-assisted extraction of pectin from artichoke by-products. An artificial neural network approach to pectin characterization

Carlos Sabater[a], Víctor Sabater[b], Agustín Olano[a], Antonia Montilla[a] , Nieves Corzo[a]

[a] *Instituto de Investigación en Ciencias de la Alimentación CIAL, (CSIC-UAM) CEI (UAM + CSIC), C/ Nicolás Cabrera, 9, E-28049, Madrid, Spain*

[b] *Sección de Ingeniería Química, Universidad Autónoma de Madrid, 28049, Madrid, Spain*

Link: https://www.sciencedirect.com/science/article/pii/S0268005X1930075X

A B S T R A C T

Artichoke (*Cynara scolymus* L.) by-products can be used as a good source of pectin. The aim of this work was to compare different pectin extraction methods: power ultrasound (US), enzymes, combination of US and enzymes (US + E), and acids (nitric and sodium citrate). After 6 h, pectin yield was higher when US was applied in combination with Celluclast®1.5 L (up to 13.9%). Structural characterization showed that US-extracted pectins had lower weight-average molecular weight (M_w) values (146–155 kDa) than pectin extracted with US + E (160–267 kDa) and acid-extracted pectin (329–352 kDa). Monomeric composition

reflected that pectin extracted with acids had the highest galacturonic acid (GalA) contents (82.2–90.2%) and the lowest degree of branching [Rha/GalA] (0.026–0.031). Structural characteristics of the different pectins were modelled using two artificial neural networks (ANN) considering composition parameters (Model I) and pectin FT-IR spectra (Model II). In addition, a third ANN was built to determine differences and similarities in the GC-MS spectra of monomeric composition (identified and unidentified monosaccharides) (Model III) showing characteristic patterns with high accuracy rates (above 95% on the test set). Structural differences depending on the extraction method of pectin have been established and these models could be applied to pectin from other sources

Keywords: Artichoke pectin extraction, Ultrasounds, Cellulase, Nitric acid, Sodium citrate, Artificial neural network

INTRODUCTION

Pectins are complex polysaccharides widely used as functional food ingredient because of their technological and biological properties. Usually pectin is extracted from industrial by-products such as citrus peel, apple pomace and sugar beet pulp (Adetunji et al., 2017). However, nowadays there is an overall interest in finding alternative sources of pectin which could have different structures and enhanced physicochemical and biological properties. With this aim, non-traditional sources of pectins such as cranberry, onion, garlic, banana, mango, pumpkin, peach, rapeseed, papaya (Adetunji et al., 2017) or artichoke by-products (Sabater et al., 2018a) have been studied.

The extraction process exerts a great influence on pectin structure. Chemical extraction employing strong acids such as hydrochloric, nitric and sulphuric acid and high temperature (70–100 °C) has been traditionally used in the industry (Maric et al., 2018; Adetunji et al., 2017; Yapo & Koffi, 2008). Nowadays, other food-grade extraction agents such as citrate are also studied (Marić et al., 2018; Muñoz-Almagro et al., 2018; Adetunji et al., 2017). Other extraction methods based on enzyme utilization

have been proposed as an environmentally-friendly alternative to acidic extraction with the advantage of the conservation of the original pectin structure (Marić et al., 2018). However, these methods can be highly time consuming, as observed by Sabater et al. (2018a) during pectin extraction from artichoke using Celluclast®1.5 L (Celluclast), a commercial cellulase preparation.

One promising and innovative extraction technique includes the assistance of power ultrasound (US) aimed at improving the yield, decreasing the energy consumption and shortening the processing time (Adetunji et al., 2017). The enhancement of extraction is mainly attributed to acoustic cavitation, generating microjets, shear forces and shockwaves at lower frequencies (20–100 kHz). Cavitation creates points with very high temperatures (which increase solubility and diffusivity) and pressures (which facilitate penetration and mass transfer) at the interface between a liquid medium and solid matrix subjected to ultrasonication; in addition, the microjetting and microstreaming effects cause the disintegration of solid materials and the disruption of cell walls. This phenomenon involves an increase in the contact between the solvent and the cell contents, enhancing mass transfer and increasing the extractive power (Soria et al., 2017; Chandrapala et al., 2013). As a result, the combination of US and enzymatic treatment can provide potential benefits. In general, it has been observed that, despite the reduction of the enzyme activity, the outcome of the combined process was positive (Nadar et al., 2018; Soria et al., 2017; Szabó & Csiszár, 2013). Numerous authors studied procedures for ultrasound-assisted enzymatic extraction (UAEE) of biomolecules obtaining positive results. Thus, Wang et al. (2014b) extracted arabinoxylan, finding that the extraction yield noticeably increased with US power from 50 W (6%) to 200 W (12%), while power above 300 W resulted in a lower yield (6%). Similar results were obtained by Liu et al. (2014) during UAEE of polysaccharides from *Lycium barbarum*, obtaining under optimum conditions (20 min) a yield of 6.31%, which was almost similar to the yield of 6.81% obtained by enzyme-assisted extraction but much more effective due to the extraction time needed (91 min) at the same conditions. Similarly, Xu et al. (2015) applied this combined technology to optimize the extraction of polysaccharides from blackcurrant. The influence of US waves on the activity and stability of enzymes has been shown to be specific for each

enzyme and dependent on sonication parameters (Rico-Rodríguezet al., 2018; Delgado-Povedano & de Castro, 2015).

On the other hand, several methodologies including multivariate data analysis may be used to study differences in pectin structure and composition. In general, data modelling allows valuable information to be discovered on chemical/structural properties and biological events (Käll et al., 2007). For this purpose, machine learning algorithms may be employed. Supervized and semi-supervized classification methods support *a priori* known data structures to train patterns and rules to predict new data. Machine learning has been widely employed in the fields of metabolomics and food science (Sabater et al., 2018b; Lin et al., 2014; Uarrota et al., 2014). Some of their applications are product characterization and component analysis (Fabris et al., 2010; Fernández Pierna et al., 2005), formulation of microparticles for polyphenol delivery (Belscak-Cvitanovic et al., 2015), detection of adulterations (Lim et al., 2017), ultrasonic extraction of phenolic compounds from wine lees (Tao et al., 2014) and study of starch gelatinization (Tao et al., 2018). In recent works, these algorithms have been used for a comprehensive characterization of pectic oligosaccharide substructures (Sabater et al., 2019a; Sabater et al., 2019b) and more studies dealing with the characterization of larger pectin structures are needed.

These algorithms can also be applied for pattern recognition using mass spectrometry (MS) data (Jha et al., 2016), the preferred analytical technique to deepen knowledge on the structural properties of pectin and pectin derivatives (Ognyanov et al., 2016). To help spectra interpretation, other computational tools assisting structure elucidation have been developed, such as *in silico* fragmentation methods (Sabater et al., 2019a; Sabater et al., 2019b; Ruttkies et al., 2016). Among them, *Competitive fragmentation modelling* (CFM) generates MS fragmentation models for several techniques including GC-EI-MS, obtaining promising results (Allen et al., 2016; Allen et al., 2015). Therefore, the aim of this work was to compare different pectin extraction methods, power ultrasound (US), enzymes, combination of US and enzymes (US + E), and acids (nitric and sodium citrate) from artichoke by-products. A structural characterization using artificial neural networks (ANNs) to assess I) differences in pectin structural parameters (Model I), II) differences in pectin FT-IR spectra (Model II)

and III) differences and similarities in the MS spectra of monomeric composition (identified and unidentified monosaccharides) of the various extracted pectins (Model III) has been carried out.

MATERIALS AND METHODS

Standards and samples

Analytical reference substances such as fructose, $_{D}$-xylose, $_{D}$-arabinose, $_{L}$-rhamnose, $_{D}$-galactose, $_{D}$-mannose, $_{D}$-glucose, galacturonic acid (GalA), and β-phenyl glucoside were purchased from Sigma Aldrich (Steinheim, Germany). Commercial multienzymatic preparations Celluclast®1.5 L (Celluclast, a cellulase from *Trichoderma reesei*) and Viscozyme®L (endo-1,3 (4)-β-glucanase from *Aspergillus aculeatus*), were generous gifts from Novozymes (Bagsvaerd, Denmark). Cellulase and pectinase activities were measured using the dinitrosalicylic acid (DNS) method. Industrial artichoke (*Cynara scolymus* L.) by-products were kindly supplied by Riberebro® (La Rioja, Spain).

Pectin extraction from artichoke by-products

Several extraction methods of pectin from artichoke by-products were used. All extraction experiences were made at least by duplicate.

Enzymatic extraction

Extractions were carried out in 50 mL polypropylene tubes with 15 mL of artichoke by-products suspension (6.5% w/v) prepared in sodium acetate buffer 50 mM pH 5.0 incubated with 10.1 U g^{-1} of Celluclast, with constant shaking (200 rpm), at 50 °C during 6 h. Pectin was purified according to Sabater et al. (2018a).

Ultrasound-assisted extraction

US treatments of artichoke by-product were carried out with Celluclast (10.1 U g^{-1}) (US + E) and without enzyme (US). The ultrasonic processor (450 Digital Sonifier, Branson Ultrasonics Corporation, Danbury, CT, USA) operated at a frequency of 20 kHz, with maximum power value of 400 W. A probe with a microtip horn of 3 mm diameter was immersed 2 cm with respect to the liquid surface into a 50 mL polypropylene tube (3 cm diameter, 11.5 cm height) with 15 mL of artichoke by-products suspension (3.25% w/v for US and 6.5% w/v for US + E treatments) suspended in sodium acetate buffer 50 mM pH 5.0. Pulsed US (2 s on/1 s off, US intensity value 81.7 W cm^{-2} corresponding to 30% amplitude) was the operating mode (Muñoz-Almagro et al., 2017). The temperature (50 ± 0.1 °C) was registered with a temperature sensor (error ± 0.1 °C) and kept constant using an ice-water bath. The extraction processes were conducted with constant shaking (200 rpm) and several extraction times were studied (1, 2, 4 and 6 h). After US extraction, pectin was purified following the method of Sabater et al. (2018a).

Acid extraction

To extract pectin from artichoke by-products using acids (citric and nitric) the method of Muñoz-Almagro et al. (2018) for pectin extraction from sunflower by-products was followed.

Sodium citrate. 2 g of artichoke by-product were treated with sodium citrate 0.74% (5% w/v solids) at 72 °C for 3.23 h at constant shaking (150 rpm). The pH was adjusted to 3.25 with citric acid 2.5 M. The mixture was cooled in an ice-water bath and centrifuged at 3700*g* for 10 min. The precipitate was washed with two vol of acidified ethanol (0.2% HCl, v/v) and then centrifuged. The supernatants were collected, mixed and precipitated with two volumes of 96% ethanol (0.2% HCl v/v) and kept overnight at 4 °C. The solution was centrifuged, the supernatant discarded and the pellet washed with acidified ethanol (0.04% HCl, v/v; pellet/solvent ratio 1:2), centrifuged again and washed with ethanol 96% (pellet/solvent ratio 1:2), centrifuged and the drained precipitate was freeze-dried.

Nitric acid. 2 g of artichoke by-product were treated with nitric acid 0.24% (5% w/v solids, pH 2) at 90 °C for 3.95 h at constant shaking (150 rpm). The mixture was cooled in an ice-water bath and purified as previously described for citrate extraction.

Analytical techniques

FT-IR spectra

In order to record pectin FT-IR spectra and to determine the degree of methyl-esterification (DM), freeze-dried samples of pectin extracted by the studied methods were analysed by FT-IR (Sabater et al., 2018a).

High performance size-exclusion chromatography with evaporative light scattering detector (HPSEC-ELSD)

Molecular weight (M_w) of extracted pectin was determined by HPSEC-ELSD following the method described by Sabater et al. (2019b). The separation of carbohydrates was carried out on a TSK-GEL G5000PWXL column (300 mm × 7.8 mm, 10 μm particle size, M_w exclusion limit 1000 kDa) and TSK-GEL G2500PWXL column (300 mm × 7.8 mm, 6 μm particle size). V_0 (void volume) and V_T (total volume) of chromatographic system were 9.6 and 28.6 mL, respectively. M_w of carbohydrates was calculated by the external calibration method using solutions of commercial pullulan standards (M_w 0.342–788 kDa) (Fluka Analytical). Weight average M_w of pectin was calculated as the weighted arithmetic mean of M_w of pectin fragments, considering their abundance (percentage): [Σ(M_w x abundance)/Σ(abundance)].

Gas chromatography coupled with mass spectrometry (GC-MS)

Monomeric composition of the different extracted pectins was studied by GC-MS. First, pectins (2% w/v in 0.05 M sodium acetate buffer; pH 5.0) were hydrolysed with 90 U mL^{-1} of Viscozyme®L preparation (Sabater et al., 2018a). Then, trimethyl silylated oximes (TMSO) of sugars released were analysed as previously described by Sabater et al. (2019b). With monosaccharide data different ratios were calculated

showing primary structural properties of pectin molecules, degree of branching (Rha/GalA), linearity pectin backbone [GalA/(Rha + Ara + Gal)] and extent of branching of RG-I [(Ara + Gal)/Rha] (Yuliarti et al., 2015; Wang et al., 2015).

Data analysis

ANOVA and Tukey's test for $p < 0.05$ were applied to all data generated.

Study of pectin composition parameters (Model I). Differences in extraction yield (%) and compositional parameters were used to classify pectin according to the extraction method through an artificial neural network (ANN) (Table R2.1, Model I). The ANN chosen was a multilayer perceptron (MLP) built with 1 hidden layer consisting in 7 neurons. The activation function was logistic. The architecture of MLP is shown in Figure R2.1.

Study of pectin FT-IR spectra (Model II). A second MLP was computed to classify pectin according to the extraction method: using 709 wavenumbers in the range of 682–3600 cm^{-1} as inputs. These wavenumbers showed significantly different ($p < 0.05$) intensities among groups and corresponded to functional groups present in pectin structure. Before spectra classification, discrete wavelet transform (DWT) was applied to FT-IR data to increase MLP classification performance. A 14 level DWT was computed (indicating the depth of the decomposition) with a “la8” decomposition filter (Daubechies orthonormal compactly supported wavelet of length = 8). In the threshold step, DWT coefficients under percentile 15% were removed to denoise the signal (Sabater et al., 2019b). The MLP used for this second study had 1 hidden layer consisting in 27 neurons.

Table R2.1. Hydrolysates of artichoke pectin extracted with different methods classified considering different parmeters using artificial neural network (ANN) study (Model I). Rha: rhamnose, GalA: galacturonic acid, Ara: arabinose, Gal: galactose, Xyl: xylose.

Extraction method	**Number of hydrolysates**
Ultrasound (US)	8
(1, 2, 4 and 6 h)	(2 per extraction)
Ultrasound and enzyme (US+E)	8
(1, 2, 4 and 6 h)	(2 per extraction)
Enzyme (E)	4
(6, 48* h)	(4 per extraction)
Sodium citrate (CIT)	4
(3.23 h)	(2 per extraction)
Nitric acid (NIT)	4
(3.95 h)	(2 per extraction)
Parameters studied in extracted artichoke pectins by ANN (Model 1)	
- Extraction yield (%)	
- Weight average molecular weight (M_w, kDa)	
- Galacturonic acid content (% total identified monosaccharides)	
- Pectic neutral sugar content (xylose, arabinose, rhamnose and galactose) (% total identified monosaccharides)	
- Non pectic monosaccharide content (fructose, mannose and glucose) (% total identified monosaccharides)	
- Degree of branching (Rha/GalA)	
- Linearity pectin backbone [GalA/(Rha+Ara+Gal+Xyl)]	
- Extent of branching of rhamnogalacturonan I (RG-I) [(Ara+Gal)/Rha]	
- Unidentified monosaccharide content (mg 100 mg^{-1})	
- Degree of methyl-esterification (DM)	

***** Sample obtained by Sabater et al. (2018a).

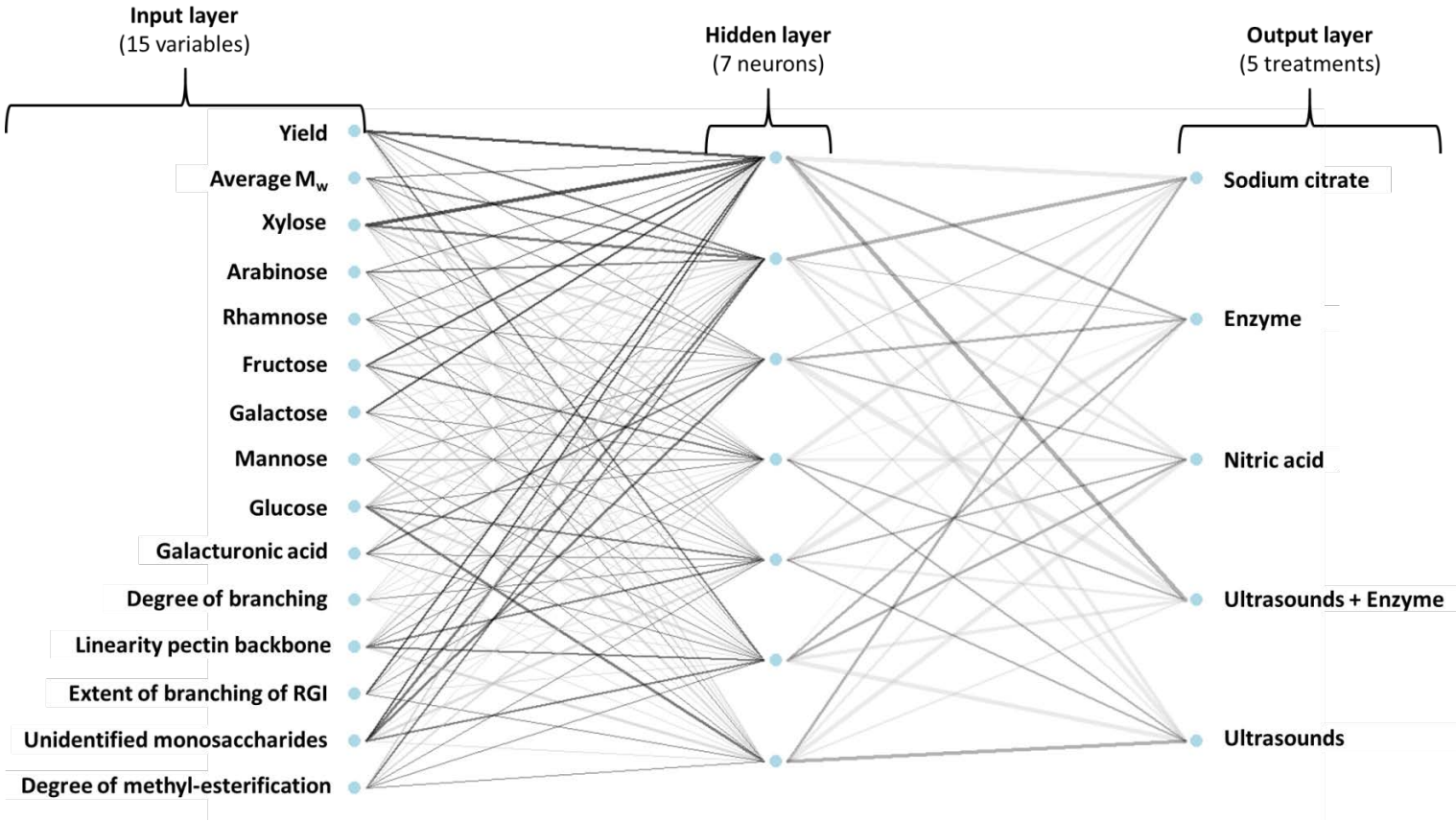


Figure R2.1. Multilayer perceptron (MLP) architecture used for studying pectin structural parameters (Model I). Fifteen variables were selected as inputs. This model classified pectin samples according to their extraction method (acids, enzymatic, ultrasounds, and combination of ultrasounds and enzymatic treatments) as output. Weights are color-coded by sign (black +, grey -) and thickness is in proportion to magnitude. RG-I: rhamnogalacturonan I.

Study of MS spectra of identified and unidentified monosaccharides present in pectin structures (Model III). A third study (Model III) comparing mass spectra of both identified and unidentified TMSO from sugars present in pectin and other polysaccharides structures co-extracted during these treatments through a third MLP was carried out (Table R2.2). Only unidentified peaks (U) released during pectin hydrolysis which were not present in enzyme preparations were considered for the study. 75 fragments in the range of *m/z* 107–540 whose abundances were statistically different among groups ($p < 0.05$) were selected. These ions might correspond to monosaccharide fragments, assessed by the competitive fragmentation modelling source code (CFM-ID) developed by Allen et al. (2016). With this aim, complete feasible fragments for candidate molecules (Sabater et al., 2019b) which could be present in pectin hydrolysates were calculated: identified monosaccharides, methyl-esterified GalA, acetylated GalA in O-2 or O-3, apiose, 3-deoxy-$_D$-manno-octulosonic acid (KDO) and two phenolic compounds, coumaric and ferulic acid. Some RG-Is (e.g., from Chenopodiaceae species such as spinach and sugar beet) are esterified with phenolics, including ferulic and/or coumaric acid, e.g., feruloylation at O-2, O-3, and O-5 of arabinose in α-L-arabinans and on O-6 of galactose in β-D-galactans (Atmodjo et al., 2013). Ferulic and *p*-coumaric acids were considered because they are present in artichoke heads (Lattanzio & van Sumere, 1987).

Before MS spectra classification, a 14 level DWT was computed. The MLP used for this third study had 1 hidden layer consisting in 57 neurons. Figure R2.2 shows its architecture. Variables were scaled and centered before computing the analyses. All the models were trained with 70% of the data, 10-fold cross-validated and then tested with 30% of data from each class (corresponding to new samples). A variable importance analysis was carried out to determine the most influential parameters in the model. For this purpose, the sum of the product of raw input-hidden, hidden-output connection weights was calculated.

Table R2.2. Number of mass spectra of identified and unidentified monosaccharides (total n = 555) obtained from GC-EI-MS analysis of hydrolysates from artichoke pectins extracted following different methodologies included in the artificial neural network (ANN) study (Model III).

Compounds	**Hydrolysates**	**Number of considered spectra**	**Number of samples**	**Total number of GC-MS spectra**
Galacturonic acid	US, US+E, E, CIT, NIT	2	28 + 1 standard	58
Arabinose	US, US+E, E, CIT, NIT	1*	28 + 1 "	29
Fructose	US, US+E, E, CIT, NIT	2	28 + 1 "	58
Galactose	US, US+E, E, CIT, NIT	1	28 + 1 "	29
Rhamnose	US, US+E, E, CIT, NIT	2	28 + 1 "	58
Xylose	US+E, E	1	12 + 1 "	13
Mannose	US, US+E, E, CIT, NIT	1	28 + 1 "	29
Glucose	US, US+E, E, CIT, NIT	1	28 + 1 "	29
Unidentified monosaccharides (U)				
1	US, US+E, E, CIT, NIT	1	28	28
2	US, US+E, E, CIT, NIT	1	28	28
3	US, US+E, E, CIT, NIT	1	28	28
4	US, US+E, E, CIT, NIT	1	28	28
5	US, US+E, E, CIT, NIT	1	28	28
6	US, US+E, E, CIT, NIT	1	28	28
7	US, US+E, E, CIT, NIT	1	28	28
8	US, US+E, E, CIT, NIT	1	28	28
9	NIT	1	4	4
10	US	1	8	8
11	US	1	8	8
12	US+E	1	8	8

* Only 1 isomer of TMSO derivatives.

US: ultrasound extraction, US+E: extraction combining ultrasounds and Celluclast, E: enzymatic extraction with Celluclast, CIT: extraction with sodium citrate, NIT: extraction with nitric acid.

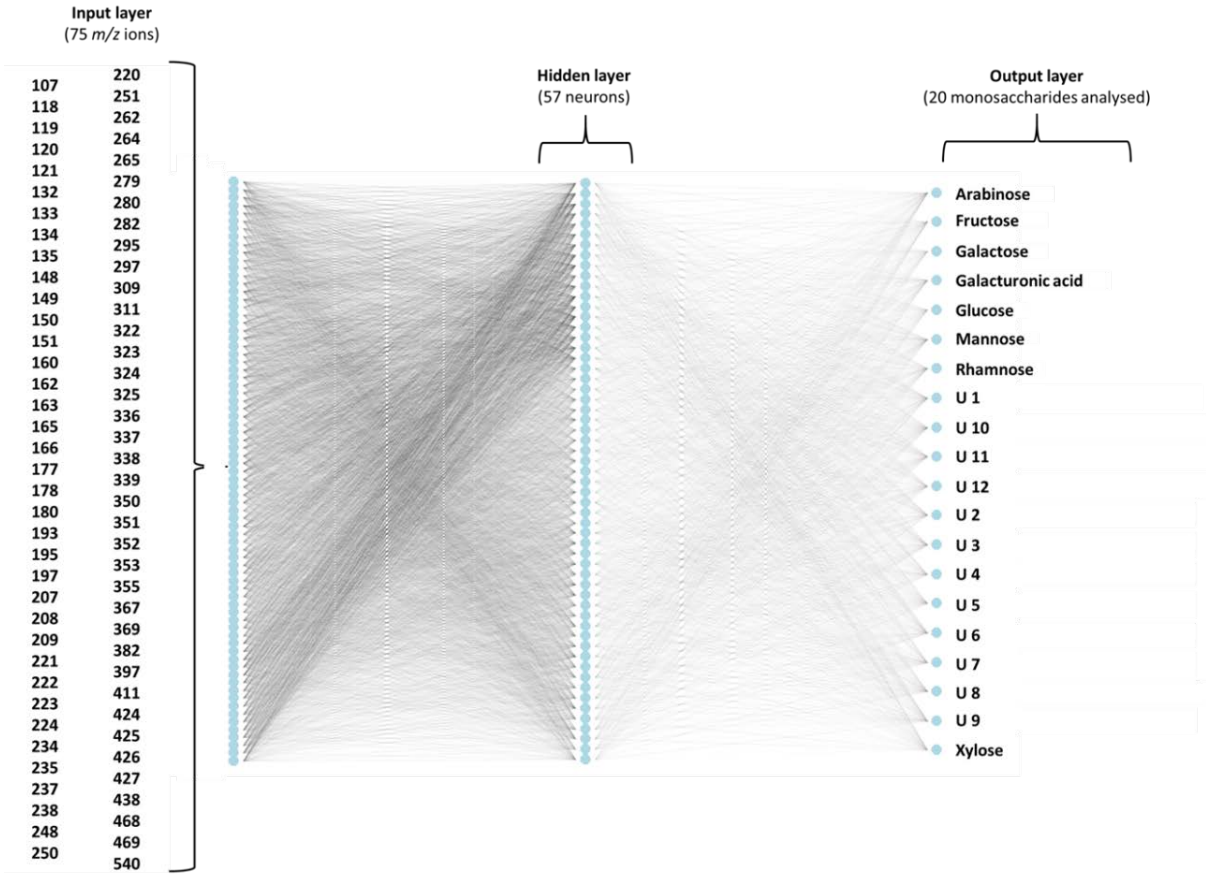


Figure R2.2. Multilayer perceptron (MLP) architecture used for studying GC-MS spectra of identified and unidentified monosaccharides found in pectin hydrolysates (Model III). 75 ions in the range *m/z* 107–540 were selected as inputs. This model classified monosaccharide GC-MS spectra according to their chemical nature (8 identified and 12 unidentified monosaccharides, U) as output. Weights are color-coded by sign (black +, grey -) and thickness is in proportion to magnitude.

The three MLP developed were compared to two simpler models, Naïve Bayes classifier (NB) and Generalized Linear Model Elastic-net (GLMNET). NB is a simple classifier based on the Bayes Theorem that assumes that the presence of a particular feature (i.e. composition parameter, *m/z* ion) in a class (i.e. extraction method, GC-MS peak) is unrelated to the presence of any other feature. Considering these features, NB uses Naïve Bayesian equation to calculate the probability for each class. The class with the highest probability is the outcome of prediction. In these models the bandwidth value (smoothing parameter) was set at 1. On the other hand, GLMNET involves a linear regression model that could be generalized (i.e. the response variable may follow different distributions than Normal distribution). Regularization methods are used to reduce possible overfitting of Generalized Linear Models and to reduce variance of the prediction error. Elastic-net regularization is based on two parameters, alpha and lambda. Alpha (comprized between 0 and 1) is used to optimize the model and it indicates the combination of two different regularization techniques, L1 and L2. All these methods try to penalize the Beta coefficients of the regression to get the important variables. Lambda is the regularization parameter or penalty coefficient, and allows adjusting the prediction error. Alpha values were set at 0.1 for all three GLMNET models while lambda was 0.028, 0.000 and 0.010 for Models I, II and III, respectively. All statistical analyses were computed on R v3.5.0. DWT was computed using wavelets package (Aldrich, 2013). MLP models were built using the RSNNS package (Bergmeir & Benitez, 2012). NB models were computed with naivebayes package (Majka, 2019). GLMNET models were built using glmnet package (Friedman et al., 2010).

RESULTS AND DISCUSSION

The yield of pectin extracted from artichoke by-products following different methodologies is shown in Table R2.3. As can be seen lower pectin yield (7.1%) was obtained when US was applied to artichoke by-products compared to enzymatic extraction (11.3%) over 6 h. However, the highest extraction yield was achieved when US + E were used (13.9%). In contrast, conventional pectin extraction showed the lowest yields (4.8 and 5.8% for sodium citrate and nitric acid, respectively). The same

trend was obtained by Li et al. (2014) who achieved the highest yield of water-soluble dietary fibre from apple pomace when they used enzymes and the lowest yield when acids were used, followed by US. With respect to conventional acid extraction, nitric acid has also been more efficient to extract pectin from sunflower heads (Muñoz-Almagro et al., 2018).

Numerous studies reported that US application enhances yields of polysaccharide extraction in general (Delgado-Povedano & de Castro, 2015; Ebringerová & Hromádková, 2010) and pectin extraction, in particular, i.e. from apple pomace (Panchev et al., 1994), passion fruit peel (Dos Santos et al., 2017), grapefruit peels (Wang et al., 2017), and sisal waste (Yang et al., 2018b).

Characterization of pectin extracted from artichoke by-products using different methods

Considering that enzymatic extraction of artichoke pectin using Celluclast preparation requires long times (48 h) (Sabater et al., 2018a), US extractions, with and without enzyme, were studied as an alternative method to reduce the extraction time. In addition, a comparison with conventional extraction methodologies using enzyme, nitric acid and sodium citrate has been carried out. Characterization of extracted pectins, including yield, weight average M_w, monomeric composition and related parameters, and DM, is shown in Table R2.3.

Monomeric composition. Identified (n = 8) and unidentified (U, n = 12) monomeric compounds were found in GC-MS of hydrolysates of artichoke pectin extracted with the different methods assayed (Supplementary Material Figure S2.1); quantitative data are compiled in Table R2.3. Pectin obtained with acids showed the highest GalA contents (90.2%), especially pectin extracted with nitric acid, similar to the results obtained by Muñoz-Almagro et al. (2018) to extract pectin from sunflower by-products (85.4–88.9%). Consequently, these pectins showed lower neutral sugar contents. On the contrary, low GalA contents were found in pectin extracted with US (61.6–68.3%) and especially in US + E treatments (59.3–66.4%). In both cases, GalA content decreased with US extraction time and after 6 h of the process, it was lower than 65%, the limit to consider pectin as a commercial food additive (E-440) (Morris et al., 2013). Regarding

pectic neutral sugars, rhamnose ranged from 1.8 to 3.3% while galactose and arabinose contents were higher in pectins treated with US and US with enzyme for prolonged times (6.2–6.9% and 15.7–17.9%, respectively). Xylose was found only in pectin enzymatically extracted (0.47%) and pectin treated with US and enzyme (0.5–1.6%). This pentose may be part of pectin, constituting the xylogalacturonan domain (Mohnen, 2008), or of hemicellulose, as xylan (Peng et al., 2012); it is possible that the xylanase activity of Celluclast favours the release of xylan and its presence in the extracted polysaccharide (Berlin et al., 2007). With respect to non-pectic monosaccharides, glucose, mannose and fructose were high when US was applied reaching up to 11.5, 1.61 and 0.75% respectively. These results highlight the fact that US extraction processes are less selective and other polysaccharides like cellulose, hemicellulose and fructans may be extracted, increasing their presence with extraction time. This lack of selectivity was partially offset by the US + E combination, however cellulase activity was not enough to hydrolyse all cellulose. This could be due to a loss of cellulase activity. Szabó and Csiszár (2013) reported that US power greatly affect the cellulase activity of Celluclast, so a treatment with probe (40 kHz, 500 W, 65 min) produced 12 and 25% of enzyme activity loss when amplitude was 40 and 80% respectively. However, the outcome of the enzyme-US combination was always positive. The advantageous effects of sonication on the heterogeneous systems are attributed to the increase in the mass transfer, due to acoustic cavitation, generating microjets, a strong physical agitation at the interface between a liquid medium and solid matrix (Soria et al., 2017). Also with Celluclast, Nguyen and Le (2013) observed that changes in ultrasonic intensity and sonication time significantly affected enzyme activity, increasing cellulolytic activity when ultrasonic intensity was increased from 0 to 6 W/mL and treatment time of 80 s, using a horn at 20 kHz, however the times were very short.

Table R2.3. Yield (mg of extracted pectin / 100 mg dry matter of artichoke by-product powder**;** %) and characterization of artichoke pectin extracted using different methodologies: ultrasounds (US), combination of ultrasounds and Celluclast (US+E), Celluclast (E), sodium citrate and nitric acid.

Treatments (h)	Yield (%)	Average M_w (kDa)	Monomeric composition (% total identified monosaccharides)								Degree of branching Rha/GalA	Linearity pectin backbone GalA/(Rha+Ara+Xyl+Gal)	Extent of branching (Ara+Gal)Rha	Unidentified Mn (mg 100 mg^{-1} Dry Matter)	DM (%)
			GalA	Xyl	Ara	Rha	Fru	Gal	Man	Glc					
Sodium citrate 3.23	4.8^{e} (1.3)*	352^{a} (8)	82.19^{b} (2.54)	-	8.57^{d} (0.96)	1.75^{c} (0.54)	0.29^{g} (0.02)	2.75^{f} (0.16)	$0.58^{c,d}$ (0.26)	5.03^{e} (0.29)	0.026^{d} (0.003)	5.85^{b} (0.44)	5.71^{e} (0.20)	$4.37^{b,c,d}$ (0.53)	35.4^{a} (1.1)
Nitric acid 3.95	$5.8^{d,e}$ (0.4)	$329^{a,b}$ (14)	90.15^{a} (1.12)	-	0.95^{e} (0.22)	$2.91^{a,b}$ (0.32)	0.47^{f} (0.03)	$2.98^{e,f}$ (0.36)	0.44^{d} (0.10)	2.71^{f} (0.16)	$0.031^{c,d}$ (0.003)	13.17^{a} (0.08)	1.48^{f} (0.23)	$3.24^{d,e}$ (0.02)	34.5^{a} (1.8)
US 1	$6.9^{c,d,e}$ (2.0)	$155^{d,e}$ (25)	$68.27^{c,d}$ (0.80)	-	14.03^{c} (0.99)	$2.07^{b,c}$ (0.07)	$0.50^{e,f}$ (0.02)	$5.16^{b,c,d}$ (0.21)	$1.60^{a,b}$ (0.23)	$8.37^{c,d}$ (0.30)	$0.030^{c,d}$ (0.001)	$3.21^{c,d}$ (0.15)	9.28^{a} (0.69)	2.52^{e} (0.08)	20.9^{c} (0.6)
2	$6.9^{c,d,e}$ (0.2)	$152^{d,e}$ (14)	$67.85^{c,d}$ (0.56)	-	14.01^{c} (0.82)	$2.53^{a,b,c}$ (0.18)	$0.57^{c,d,e,f}$ (0.03)	$4.78^{c,d}$ (0.02)	$0.84^{a,b,c,d}$ (0.02)	$9.41^{b,c}$ (0.44)	$0.037^{b,c}$ (0.003)	$3.19^{c,d,e}$ (0.18)	$7.43^{a,b,c,d}$ (0.21)	$2.91^{d,e}$ (0.10)	22.0^{c} (1.9)
4	$7.7^{c,d,e}$ (2.3)	$153^{d,e}$ (0.2)	$64.52^{d,e,f}$ (2.07)	-	$15.08^{b,c}$ (1.26)	$3.04^{a,b}$ (0.05)	$0.63^{b,c,d}$ (0.04)	$5.32^{a,b,c,d}$ (0.09)	$0.94^{a,b,c,d}$ (0.13)	$10.46^{a,b}$ (0.59)	$0.047^{a,b}$ (0.001)	$2.76^{d,e,f}$ (0.24)	$6.72^{d,e}$ (0.55)	$3.32^{d,e}$ (0.24)	22.8^{c} (1.0)
6	$7.1^{c,d,e}$ (2.3)	146^{e} (6)	$61.58^{e,f}$ (0.85)	-	$15.71^{a,b,c}$ (0.25)	3.26^{a} (0.36)	$0.69^{a,b}$ (0.04)	$6.23^{a,b,c}$ (0.96)	$1.03^{a,b,c,d}$ (0.08)	11.51^{a} (0.74)	0.053^{a} (0.007)	$2.45^{e,f}$ (0.19)	$6.76^{c,d,e}$ (0.37)	$3.34^{d,e}$ (0.09)	29.7^{b} (1.6)
US + E 1	$10.9^{b,c}$ (0.3)	$160^{d,e}$ (6)	$66.42^{d,e}$ (0.52)	0.53^{b} (0.13)	$15.82^{a,b,c}$ (1.33)	$2.61^{a,b,c}$ (0.47)	$0.61^{b,c,d,e}$ (0.02)	$5.54^{a,b,c,d}$ (0.88)	$1.02^{a,b,c,d}$ (0.02)	7.46^{d} (0.26)	$0.039^{b,c}$ (0.007)	$2.77^{d,e,f}$ (0.13)	$8.29^{a,b,c,d}$ (1.32)	$3.46^{c,d,e}$ (0.18)	22.2^{c} (0.0)
2	$12.4^{b,c}$ (1.8)	$188^{c,d,e}$ (34)	$64.35^{d,e,f}$ (0.90)	1.47^{a} (0.61)	$16.23^{a,b,c}$ (0.05)	$2.90^{a,b}$ (0.34)	$0.68^{a,b,c}$ (0.03)	4.54^{d} (0.37)	$1.44^{a,b}$ (0.62)	$8.39^{c,d}$ (0.40)	$0.045^{a,b}$ (0.005)	$2.72^{d,e,f}$ (0.05)	$7.21^{b,c,d,e}$ (0.70)	$3.46^{c,d,e}$ (0.29)	22.2^{c} (0.8)
4	$10.3^{b,c,d}$ (1.2)	$267^{b,c}$ (59)	$62.10^{e,f}$ (0.90)	1.51^{a} (0.04)	$17.66^{a,b}$ (0.37)	$3.05^{a,b}$ (0.18)	0.76^{a} (0.04)	$4.80^{c,d}$ (0.03)	$0.80^{b,c,d}$ (0.15)	$9.32^{b,c}$ (0.53)	$0.049^{a,b}$ (0.002)	2.43^{f} (0.06)	$7.37^{b,c,d}$ (0.58)	$3.73^{c,d,e}$ (0.35)	21.1^{c} (0.6)
6	13.9^{b} (0.1)	$224^{c,d,e}$ (37)	59.28^{f} (1.62)	1.58^{a} (0.11)	17.87^{a} (1.10)	$2.82^{a,b}$ (0.16)	0.75^{a} (0.02)	6.87^{a} (0.04)	1.61^{a} (0.01)	$9.23^{b,c}$ (0.26)	$0.048^{a,b}$ (0.004)	2.15^{f} (0.15)	$8.77^{a,b}$ (0.11)	5.90^{b} (0.92)	22.8^{c} (1.7)
Enzyme 6	$11.3^{b,c}$ (1.8)	$206^{c,d,e}$ (4)	72.33^{c} (0.84)	0.47^{b} (0.00)	13.75^{c} (0.49)	$2.09^{b,c}$ (0.09)	$0.69^{a,b}$ (0.02)	$4.21^{d,e}$ (0.36)	$1.16^{a,b,c}$ (0.02)	5.30^{e} (0.62)	$0.029^{c,d}$ (0.002)	3.61^{c} (0.08)	$8.61^{a,b,c}$ (0.30)	10.37^{a} (1.75)	20.2^{c} (0.9)
Enzyme 48**	20.3^{a} (0.0)	$228^{c,d}$ (5)	$68.68^{c,d}$ (1.23)	0.53^{b} (0.19)	$15.92^{a,b,c}$ (0.03)	3.25^{a} (0.18)	$0.55^{d,e,f}$ (0.07)	$6.38^{a,b}$ (0.57)	$1.35^{a,b}$ (0.09)	3.33^{f} (0.42)	$0.047^{a,b}$ (0.003)	$2.69^{d,e,f}$ (0.13)	$6.86^{c,d,e}$ (0.20)	$5.49^{b,c}$ (0.02)	19.5^{c} (0.0)

^{a,b,c,d,e,f,g} Statistical differences between groups; Average M_w: weight average molecular weight determined by HPSEC-ELSD, GalA: galacturonic acid, Xyl: xylose; Ara: arabinose, Rha: rhamnose, Fru: fructose, Gal: galactose, Man: mannose, Glc: glucose, Unidentified Mn: unidentified monosaccharides including peaks U1-12 (Figure S1). *: standard deviation. ** Data from Sabater et al. (2018a).

Finally, in hydrolysates of pectin an important group of unidentified monosaccharides (n = 12) was also detected (Table R2.2; Figure S2.1). They were quantified together and amounts ranging from 2.5 to 10.4 mg 100 mg^{-1} dry matter were found (Table R2.3). They have been considered for the structural study of extracted polysaccharides.

Linearity and branching. Table R2.3 shows some ratios between contents of GalA and neutral pectic sugars in the purified pectins indicating possible branching along the homogalacturonan (Yuliarti et al., 2015). Considering the monomeric composition of pectin, the degree of branching (Rha/GalA), linearity pectin backbone [GalA/(Rha + Ara + Gal)] and extent of branching of RG-I [(Ara + Ga)/Rha] were in the ranges 0.026–0.053, 2.2–13.2 and 1.5–9.3, respectively. The degree of branching was higher in pectin extracted with US and US + E due to its higher neutral sugar content, and lower in Celluclast and acid-extracted pectin because of its high GalA values. This parameter increased with prolonged US time as higher amounts of neutral sugars are released, while the linearity pectin backbone evolves inversely. The biggest differences among processed samples were observed in the other two pectin backbone parameters; pectins extracted with acids present a higher linearity and lower extent of branching of RG-I with respect to the others, highlighting the structural variation of the pectins obtained and the importance of the RG-I domain. From these molecular parameters, perhaps the most important for pectins could be the degree of branching. In general, this parameter for artichoke pectin (0.026–0.053) was similar for commercial apple pectin (0.033–0.037) (Muñoz-Almagro et al., 2017) and lower than for grapefruit peel pectin, extracted with or without US (0.12–0.13) highlighting the importance of RG-I in this case (Wang et al., 2016).

FT-IR spectra. *Degree of methyl-esterification*. Figure R2.3 illustrates FT-IR spectra of artichoke pectin extracted with different treatments, acids, US + E and E. Several bands corresponding to specific functional groups were determined, showing different intensities according to the extraction method. First, a significant band showing a wave number of 3300 cm^{-1} was observed in all samples. This band corresponds to O-H bonds stretch tension. This signal was high in acid-extracted samples and low in US + E treatments. On the other hand, the characteristic signal of C-H bonds stretch tension was

found at 2920 cm^{-1} and was high in acid treatments. A signal corresponding to carboxylic acid C=O bonds stretch was reported at 1700 cm^{-1}. As expected, the highest intensities were reported in nitric-extracted pectin while intensity values found in US treatments were slightly higher than those found in the combination of US and enzymes. Carboxylate groups were found in the 1600 cm^{-1} band, with no significant differences among treatments. Other important signals were found at 1400 and 1200 cm^{-1}, corresponding to C-O-H bend deformation and carboxylic acid C-O stretch tension, respectively. In addition, several hydroxyl and ether characteristic peaks were found between 1000 and 1300 cm^{-1}. Finally, a 600-700 cm^{-1} signal was present in every sample, corresponding to the vibrational tension frequency of the pyranoid ring. This signal was high in enzymatically extracted pectin. According to FT-IR spectra from artichoke pectin obtained by different methods could be classified as low methyl-esterified pectin or pectic polysaccharide with a degree of methyl-esterification ranging from 20.2 to 35.4%, similar to other pectins from sunflowers (38–41%) and lower than those of apple and citrus pectin previously characterized in our laboratory (72–76%) (Muñoz-Almagro et al., 2018; Muñoz-Almagro et al., 2017). DM was significantly higher in those pectins extracted with acids.

Molecular weight distribution. HPSEC-ELSD profiles of pectins extracted using different methods are shown in Figure R2.4 and their quantitative data in Figure R2.5. In general, pectin had a multimodal distribution and their patterns were not modified with the time of treatment. Pectin extracted with sodium citrate and nitric acid presented a similar bimodal pattern, with fragments > 600 kDa and around 130 kDa. US + E pectins presented three main fragments of > 600, 146–170 and 9–11 kDa. These profiles were similar to those obtained with Celluclast after 6 h of extraction that presented three fragments of 587, 130 and 11 kDa, and those obtained after 48 h with three fragments of 660, 105 and 4.8 kDa (Sabater et al., 2018a). US-extracted pectins showed a monomodal pattern, with a main fragment of 150–160 kDa and a small 9–15 kDa fraction, only corresponding at 2–5% (Figure R2.5b). In general, the most common M_w distribution of pectins (from citrus, apple, sunflower) for HPSEC-ELSD analysis is monomodal (Muñoz-Almagro et al., 2018; Muñoz-Almagro et al., 2017), however some pectins from sugar beet pulp show two peaks with different M_w (Yapo & Koffi, 2008),

or from mango nectar a multimodal pattern (Huang et al., 2018). With regard to weight average M_w of pectin (Table R2.3), US-extracted pectins showed lower values (146–155 kDa) than enzymatically obtained pectin (206 kDa), US + E (160–267 kDa) and acid-extracted pectin (329–352 kDa) possibly indicating that when only US was applied, under the conditions used, the treatment was not powerful enough to remove all of the cell-wall high M_w polysaccharides. This data agrees with that obtained by Yang et al. (2018b), who explained that the higher molar mass attained from combined enzymatic-ultrasonic extraction can be explained as follows: firstly ultrasound is a relatively mild and not destructive extraction method in contrast to the acid methods. Secondly, only US did not release the larger calcium-bound pectin molecules which would be hard to extract only with this method. Interestingly, pectin extracted with acids showed higher weight average M_w values, because fragments between 2 and 20 kDa were not present. Also Wang et al. (2016) reported that US extracted pectin from grapefruit peel (20 kHz, 800 W, tip 25 mm, 2 s on/2 s off, US power density 0.41 W/mL, 67 °C over 28 min), has a lower M_w than that extracted without US (80 °C, 1.5 h), using deionized water at pH 1.5 adjusted with HCl as the extraction solvent. They explained that the US produced severe degradation. However, in our US-treated sample the fraction of low M_w was only 2–5%.

Overall, these results indicate that the application of different extraction methods using acids, US, Celluclast or US + Celluclast gave rise to pectins that were structurally different with respect to those obtained by conventional methods and with the possibility to use them as a functional ingredient and thus widen their application area.

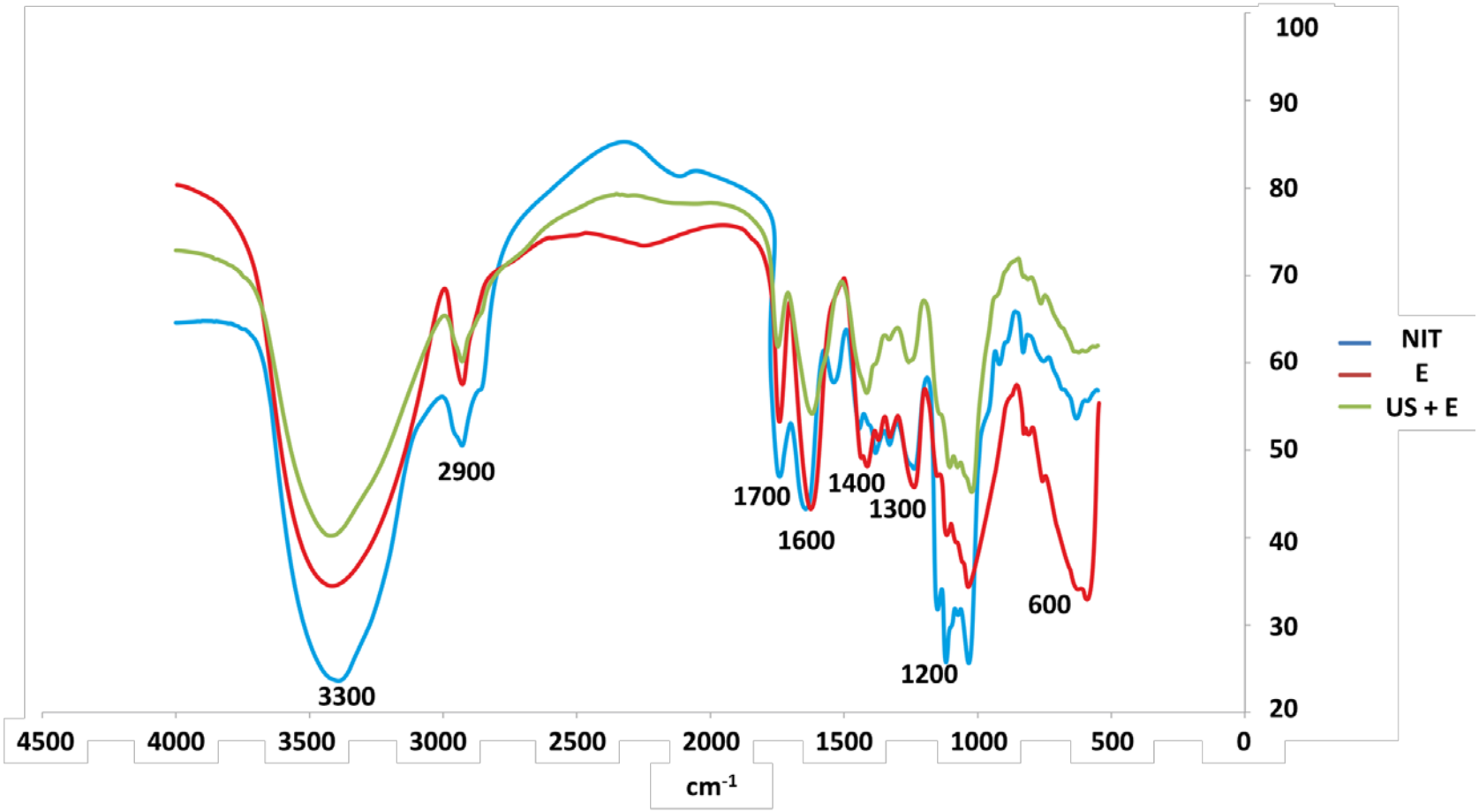


Figure R2.3. FT-IR spectra of pectin extracted from artichoke by-products by different methods. NIT: nitric acid, E: enzymatic, US + E: combination of ultrasounds and enzymes.

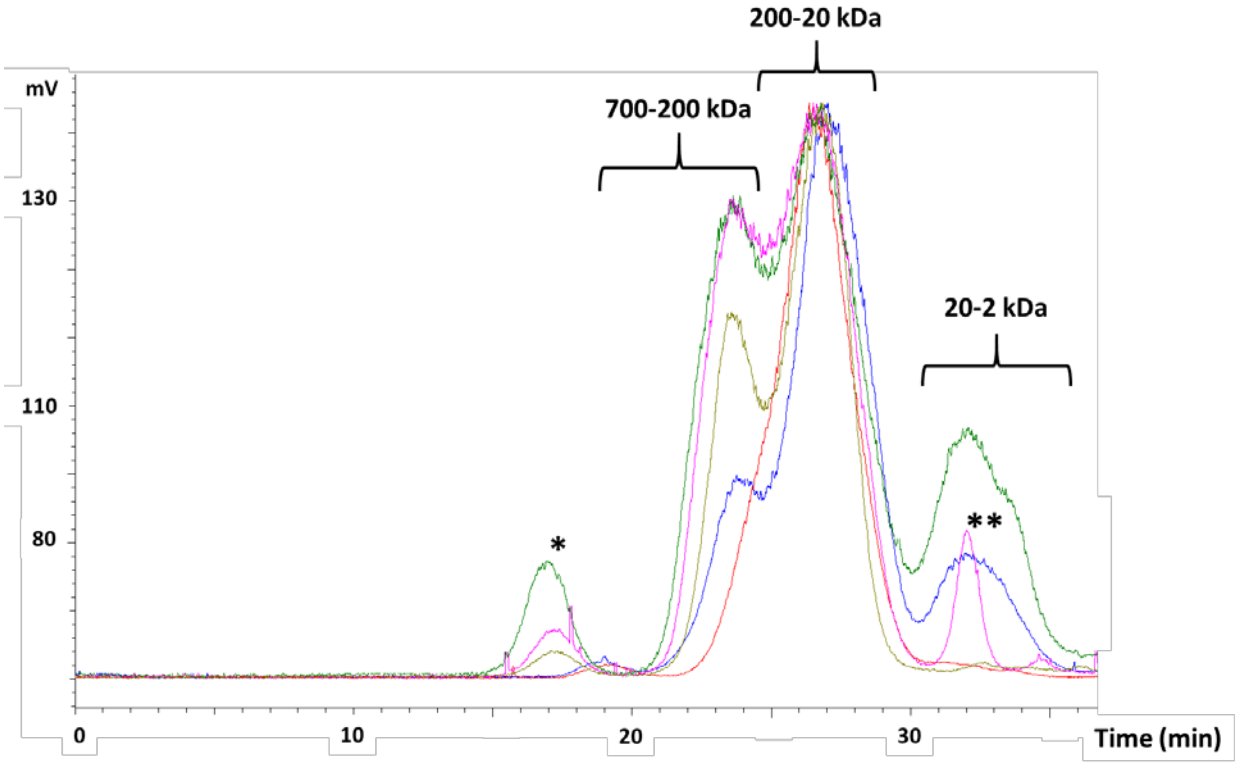


Figure R2.4. HPSEC-ELSD profiles of artichoke pectin extracted from by-products using different methods: Celluclast for 6 h (blue), ultrasounds and Celluclast for 6 h (green), ultrasounds for 6 h (red), nitric acid (brown) and sodium citrate (pink). *Compounds out of M_w range **Sodium citrate peak.

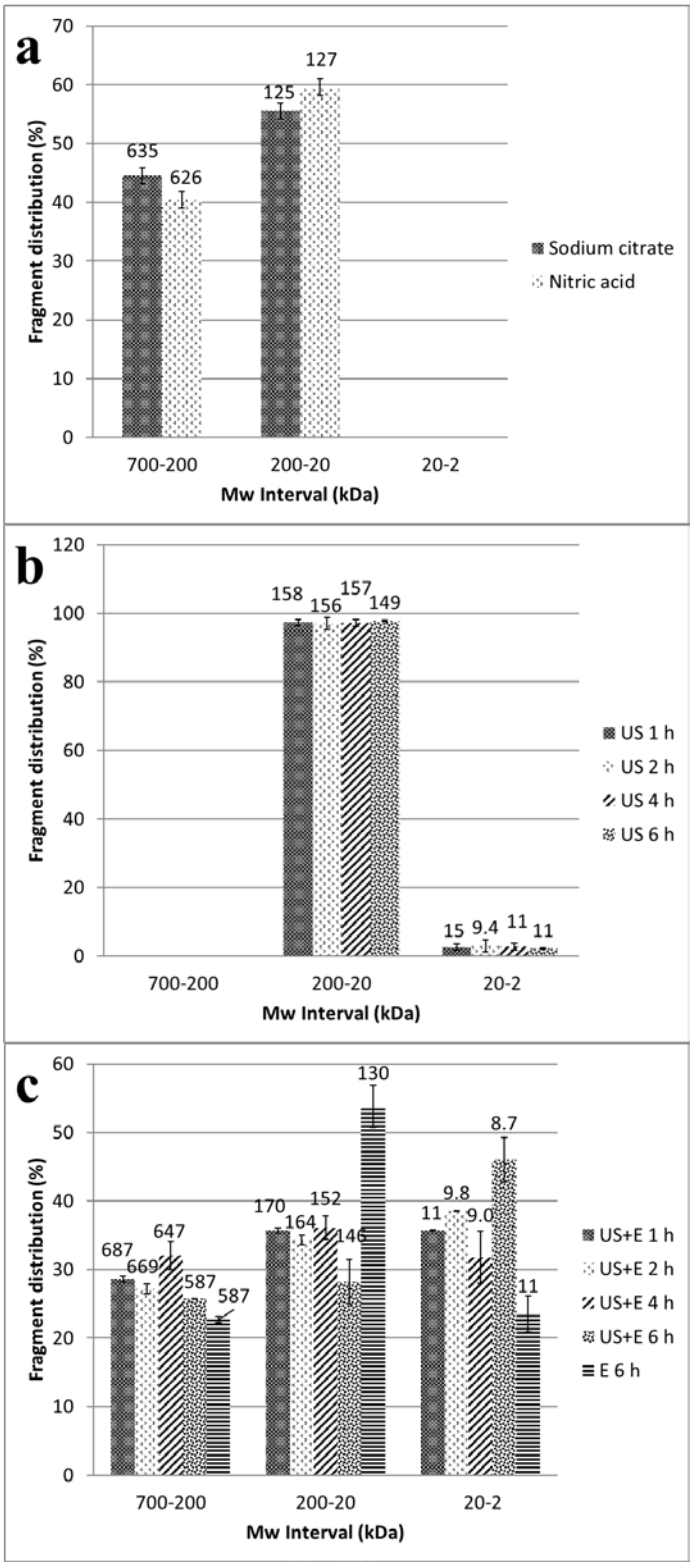


Figure R2.5. Molecular weight distribution (HPSEC-ELSD) of artichoke pectin extracted using different methods: a) sodium citrate and nitric acid, b) ultrasounds (US) and c) Celluclast (E) and combination of ultrasounds and Celluclast (US + E).

ANN modelling of pectin compositional parameters

Structural differences and the composition of artichoke pectin were confirmed through data modelling (Model I, Table R2.1). Pectin samples were classified according to their extraction method (with nitric acid, sodium citrate, Celluclast, US and US + E) considering the parameters included in Table R2.3. With this aim, an ANN was applied (Figure R2.1). This model was trained with 70% of data, 10-fold cross-validated and then tested on new samples (30% of data from each class i.e. extraction method). The training, cross-validation and test rates were 100, 91.67 and 100%, respectively. The overall kappa value was 0.86. This parameter is a more robust accuracy metric that considers the possibility of a correct classification by chance. The ANN model was compared to another two simpler models (Supplementary Figure S2.2): NB (100, 79.19 and 100% training, cross-validation and test rates) and GLMNET (100, 85.00 and 100% training, cross-validation and test rates) that showed kappa values of 0.75 and 0.80. The model performance of ANN was similar to that of GLMNET. As these values indicate, this machine learning algorithm found a reproducible classification pattern for each kind of pectin that presented high prediction rates on the independent test set, showing no misclassification in this step. No misclassification occurred during the test phase so model sensitivity, specificity, true positive and negative rates and balanced accuracy were 100%.

The variable importance of ANN was calculated to determine the most influential parameters in this classification (Supplementary Material Table S2.1). The most influential variables for modelling were yield and xylose, glucose and unidentified monosaccharides. The yield showed great importance in classifying pectin extracted with Celluclast (importance coefficient, IC 6.53), also with positive values for US + E (IC 0.57). Xylose content was important probably due to the fact that it is only present in pectins extracted with enzyme and US + E (IC 1.46 and 7.34, respectively), while in the other pectins the coefficients were negative. High glucose content resulted to be relevant for classifying pectin extracted with US (IC 7.90), and may indicate that polysaccharides other than pectin are extracted during the process whereas low glucose contents were characteristic in nitric acid (IC -5.21) and Celluclast (IC -5.03) extracted pectins. Unidentified monosaccharide content was of great importance to differentiate

artichoke pectin extracted with Celluclast (IC 5.53); this treatment produced the highest amount of these compounds. The other parameters have a minor individual relevance. In general terms, ANN modelling corroborates differences observed in experimental data and allows these products to be accurately classified according to the parameters studied. This model also highlights structural differences depending on the extraction method that may lead to different bioactivity. For example, high neutral sugar contents obtained in enzymatic treatments (including US + E), especially arabinose contents, could confer to this pectin an anticarcinogenic property (Wikiera et al., 2015a,b).

As a general overview, statistical differences among groups were calculated and they are shown in Figure R2.6(a–e) considering extraction yield (%); weight average M_w (kDa); GalA content; pectic neutral sugar (xylose, arabinose, rhamnose and galactose) content; non-pectic monosaccharide (mannose and glucose) content. Also, statistical differences between parameters related to pectin chains were included (Figure R2.6 f-h), considering degree of branching, linearity pectin backbone and extent of branching of RG-I. As has been previously indicated, enzymatic extraction (with and without US) produced the highest yields. The lowest weight average M_w values were obtained in the US extraction. Pectic neutral sugar and other monosaccharide contents were higher in US-extracted pectin, with the latter showing a less selective extraction when US was applied. Finally, pectin extracted by acids showed high GalA content and linearity pectin backbone and a lower degree of branching and extent of branching of RG-I, suggesting a less ramified structure with a different composition and different properties. The statistical tests confirm the different patterns established during the supervized classification, which have a potential application on pectin from an unknown origin considering the high accuracy rates on new samples.

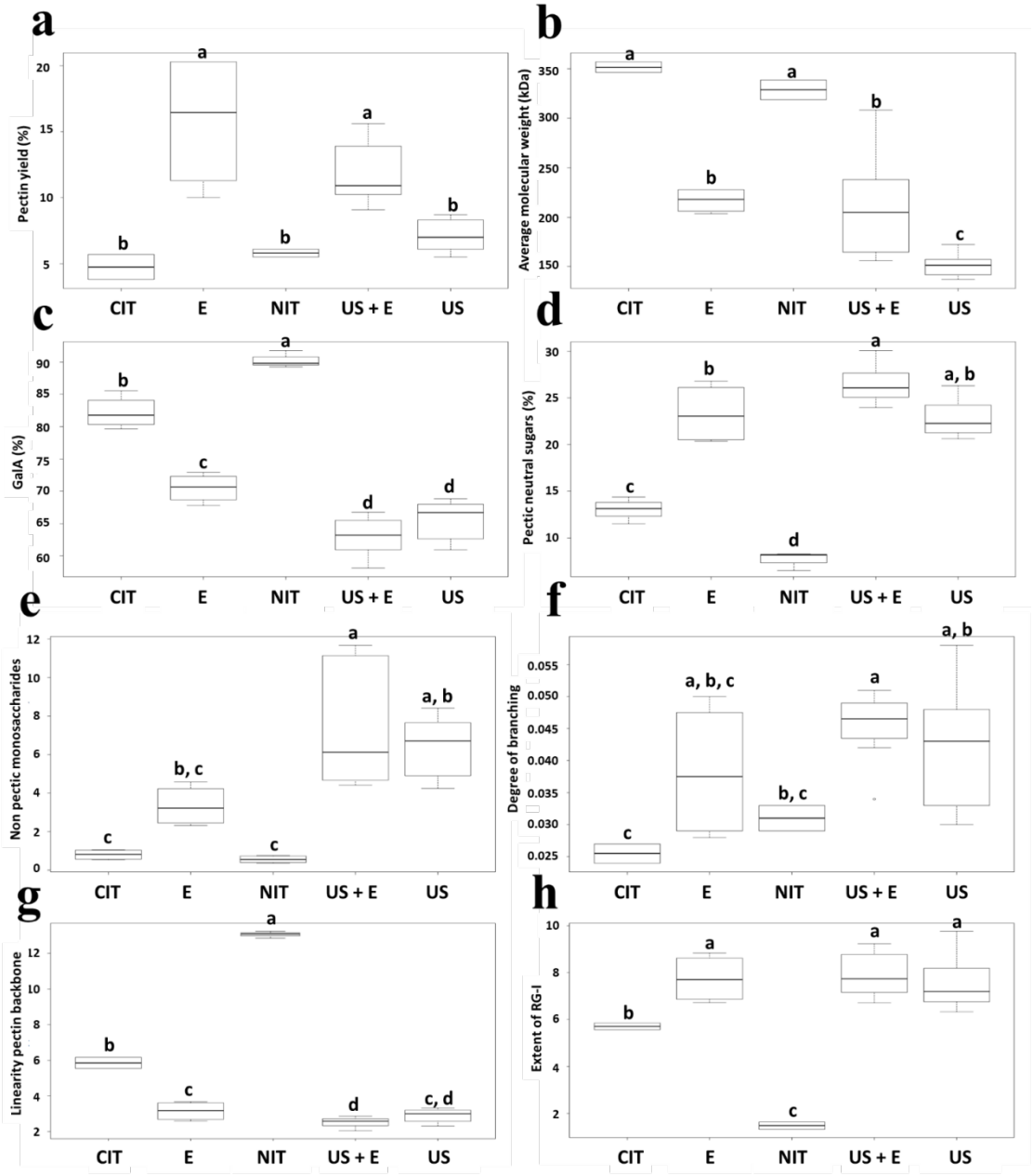


Figure R2.6. Values of pectin extraction yield (%) (a), weight average molecular weight (kDa) (b), galacturonic acid (GalA) (c), pectic neutral sugar (xylose, arabinose, rhamnose and galactose) (d) and non pectic monosaccharides (mannose and glucose) (e) contents, degree of branching (Rha/GalA) (f), linearity pectin backbone [GalA/(Rha + Ara + Ga)] (g) and extent of branching of rhamnogalacturonan I [(Ara + Ga)/Rha] (h) values, found in pectin samples extracted by different methods, US: ultrasound, US + E: combining ultrasounds and Celluclast, E: Celluclast, CIT: sodium citrate, NIT: nitric acid. [a,b,c,d] Statistically significant differences ($p < 0.05$) between pectins.

ANN modelling of pectin FT-IR spectra

Once it was demonstrated that pectin samples could be classified according to their composition parameters, these pectins were classified in a second study considering their FT-IR spectra (Model II). The training, cross-validation and test rates were 100, 83.15 and 100%, respectively. The overall kappa value was 0.82. The ANN model was also compared to another two models (Supplementary Figure S2.3): NB (100, 79.94 and 100% training, cross-validation and test rates) and GLMNET (100, 80.59 and 100% training, cross-validation and test rates) that showed kappa values of 0.71 and 0.77. The model performance of ANN was similar to that of GLMNET. However, ANN performance was slightly higher than the rest of the models although these three algorithms made no misclassifications during the test phase indicating that a highly reproducible spectral pattern was found.

Similar to the previous model (Model I), model sensitivity, specificity, true positive and negative rates and balanced accuracy were 100%, indicating its high prediction rates. The most influential wavenumbers were then determined for each treatment (Supplementary Material Table S2.2). A total of 49 most relevant wavenumbers ranging from 1000 to 2828 cm^{-1} were selected to discuss statistically significant differences ($p < 0.05$) between FT-IR spectra of pectin extracted with different treatments:

Some wavenumbers (1000, 1019, 1024, 1042, 1044, 1055, 1063, 1081, 1083, 1103, 1104, 1131, 1142, 1145 and 1154 cm^{-1}) showed a high intensity in pectin extracted with acids and Celluclast and may correspond to the characteristic bend deformation of hydroxyl and ether groups. Other wavenumbers (1258, 1271, 1308, 1680, 1684, 1718, 1723, 1745 and 1762 cm^{-1}) were low in US treatments. As previously commented, wavenumbers 1258–1308 may correspond to hydroxyl and ether groups while wavenumbers 1680–1762 indicate carboxylic acid C=O bonds stretch. Some wavenumbers (2703, 2723, 2748, 2759 and 2772 cm^{-1}) were high in Celluclast treatments. Finally, other wavenumbers (1230, 1700, 1769, 1771, 1772, 1773, 1774, 1776, 1777, 1778, 1817, 1825, 1830, 1831, 1838, 2786, 2800, 2805, 2822 and 2828 cm^{-1}) were high only in citrate and Celluclast treatments; wavenumber 1230 cm^{-1} corresponds to hydroxyl and ether groups, 1700-1838 cm^{-1} to C=O groups, and 2786–2828 correspond to O-H bonds stretch.

This FT-IR spectra study confirms structural differences of pectin depending on the extraction method and gives complementary information to the first study based on composition parameters. On the other hand, it is important to note that, if desired, pectin could be classified according to the extraction method with this simple analysis, although deep structural information would not be obtained. To achieve this purpose other structural determinations, such as those discussed in the model I, and especially a study of spectra GC-EI-MS, made in the model III are necessary.

ANN study of GC-EI-MS spectra of identified and unidentified monosaccharide constituents of artichoke pectin

To gain deeper knowledge of the structures of extracted pectin, a GC-MS study of their monomeric composition, both of identified and unidentified sugars detected in the chromatographic profiles of hydrolysates (Supplementary Material Figure S2.1) was carried out (Table R2.2, Model III). As can be observed, identified monosaccharides (xylose, arabinose, rhamnose, fructose, galactose, mannose, glucose and GalA) and 12 unidentified peaks were present in these structures. Unidentified peaks 1–8 (U1-U-8) were present in pectin extracted with all studied methods. U9 was found in nitric acid-extracted pectin while U10 and U11 were specific of pectin extracted with US and U12 was only present in pectin extracted with US + E (Table R2.2). Similarities and differences in the mass spectra of these identified and unidentified compounds were established using a third ANN (Model III). One standard for each identified monosaccharide was also included in the study, as reference. A total of 555 mass spectra were classified considering 75 to be relevant *m/z* ions (Figure R2.2). To increase model performance, each MS spectrum was decomposed and reconstructed using the DWT. Then, ANN classified each spectrum according to each identified/unidentified peak. The training, cross-validation and test rates were 99.74, 97.19 and 96.97%, respectively. The overall kappa value was 0.95.

The most influential *m/z* ions in the classification were determined (Supplementary Material Table S2.3). In addition, model sensitivity, specificity, true positive and negative and balanced accuracy rates were calculated (Supplementary Material Table S2.4). Again, MLP found a reproducible classification pattern that showed high

accuracy on the test set (above 95%). Interestingly, model sensitivity was 100% for the studied compounds with the exception of U7 (80.0%) galactose (81.8%) and rhamnose (94.1%). Moreover, specificity rates were ≥98% in all cases. In general, true positive rates were 100% with the exception of rhamnose (94.1%), U11 (66.7%) and U6 (62.5%), probably due to structural similarities with other monomeric compounds that may interfere with the model predictions. However, true negative rates were ≥98% and the lowest balanced accuracy values found were 90.0 and 90.9% for U7 and galactose, indicating in all cases a high predictive power. Similar to the previous models, ANN was compared to other models (NB and GLMNET) that reached training, cross-validation and test rates of 97.69, 89.03, 95.76% (NB) and 98.46, 86.33, 93.33% (GLMNET). Kappa values were of 0.88 and 0.86 for NB and GLMNET, respectively. Although these simpler models allow GC-MS spectra to be classified accurately showing high sensitivity and specificity (Supplementary Table S2.5 and Table S2.6), some of the compounds studied could not be classified, especially unidentified compounds U9 and U12, probably due to the complexity of its spectra. Therefore, ANN outperformed the rest of the models when tested on new samples (Supplementary Figure S2.4).

To deepen the study, possible chemical structures of studied fragments were suggested employing CFM-ID code (Allen et al., 2016). This code is able to calculate all possible fragments that can be formed from one molecule (the ones included in our *in silico* fragmentation library), including both immediate descendants and fragments formed from these descendants. Therefore, the chemical origin of relevant *m/z* ions has been determined (Supplementary Material Figure S2.5). Proposed structures of the most interesting *m/z* ions (n = 27) are shown in Table R2.4. It should be noted that these structures contain nitrogen and trimethyl silyl groups because they correspond to TMSO fragments of monomeric compounds that could be used as marker ions. In addition, non-derivatized chemical substructures that are present in the original carbohydrate molecule are included in Table R2.4. Then, differences in the GC-MS spectra of identified and unidentified compounds were summarized, considering statistically significant differences ($p < 0.05$) in their abundances (Supplementary Material Figure S2.6):

Table R2.4. Possible chemical structures of the most relevant *m/z* ions introduced as neural network inputs, determined by CFM-ID. These ions correspond to TMSO fragments from cell wall monosaccharides and cumaric and ferulic acids.

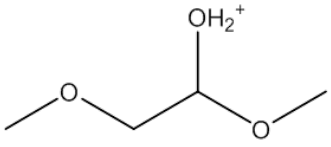
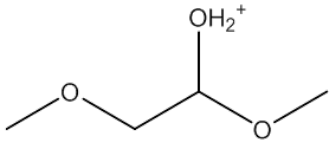
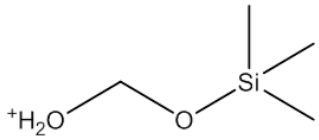
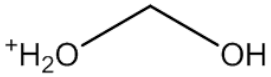
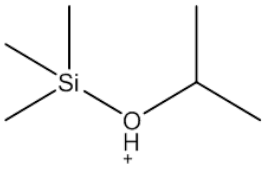
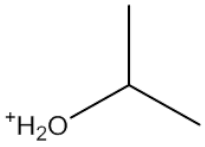
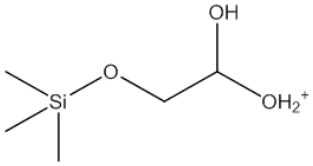
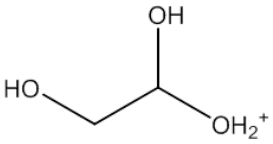
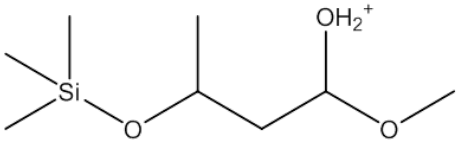
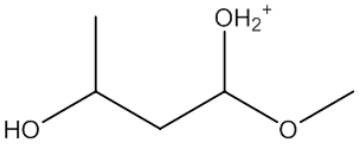
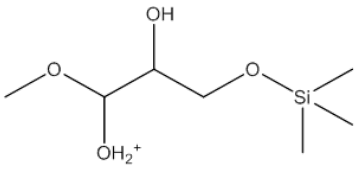
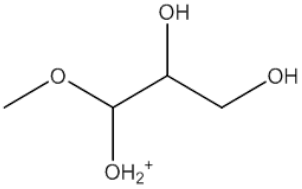
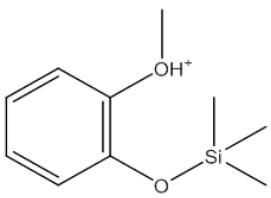
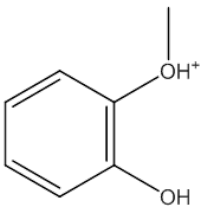
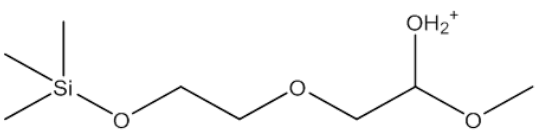
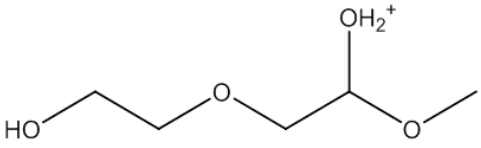
m/z	TMSO Structure	Original carbohydrate structure
107		
121		
133		
151		
193		
195		
197		
209		

Table R2.4. Cont.

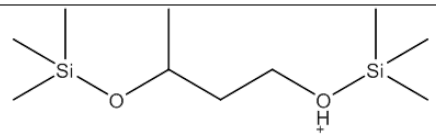
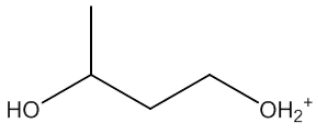
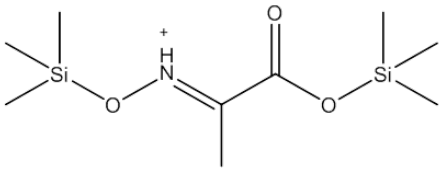
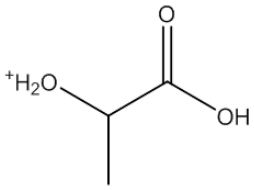
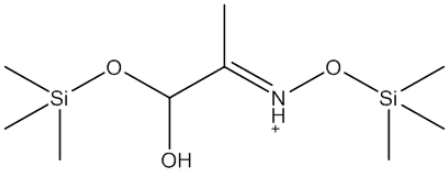
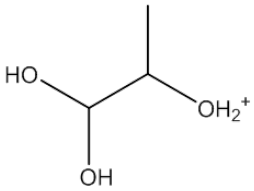
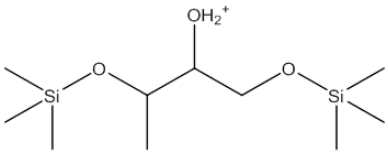
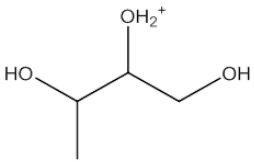
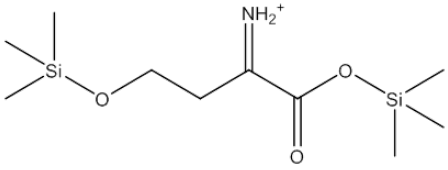
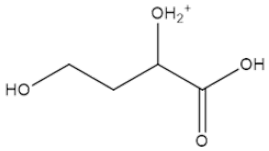
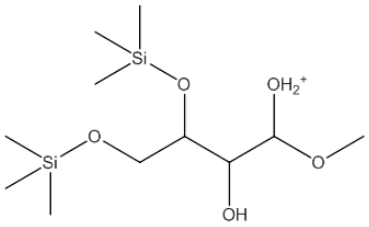
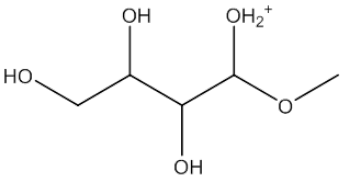
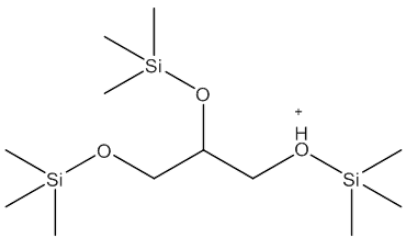
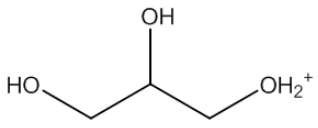
m/z	TMSO Structure	Original carbohydrate structure
235		
248		
250		
251		
262		
297		
309		

Table R2.4. Cont.

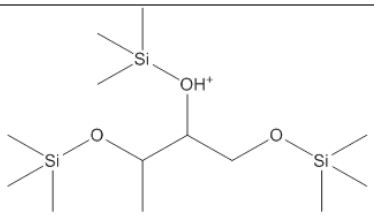
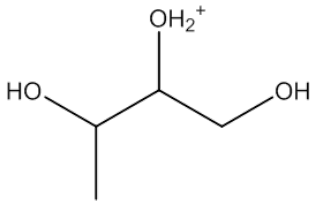
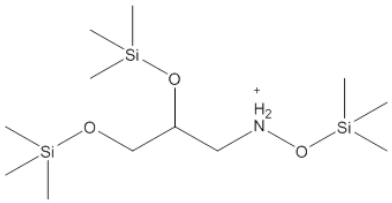
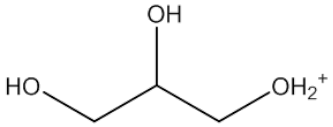
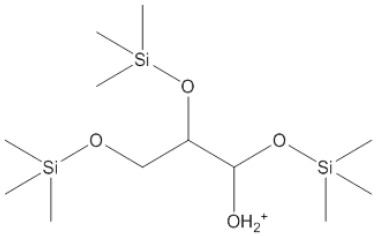
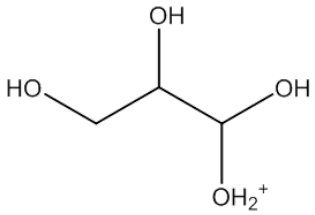
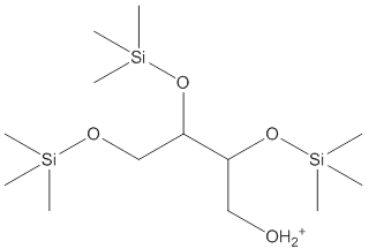
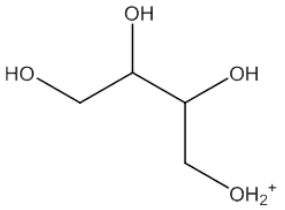
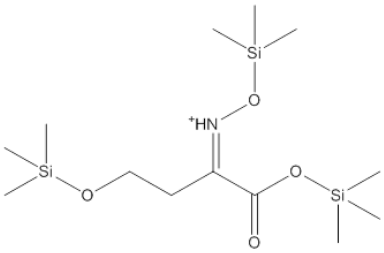
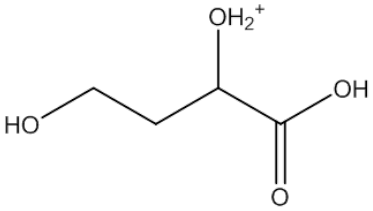
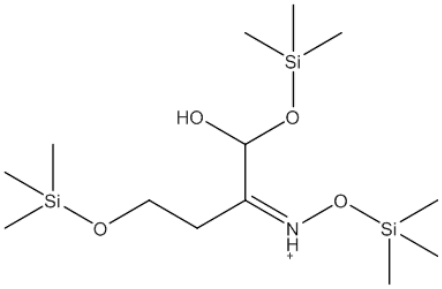
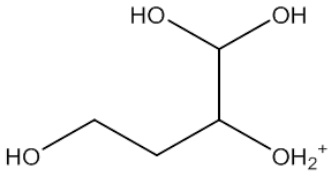
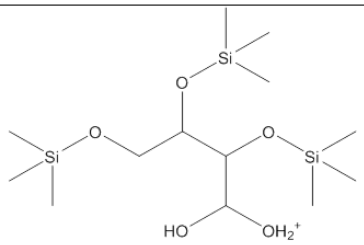
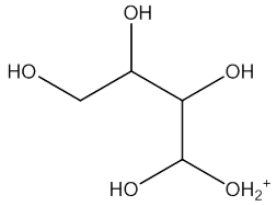
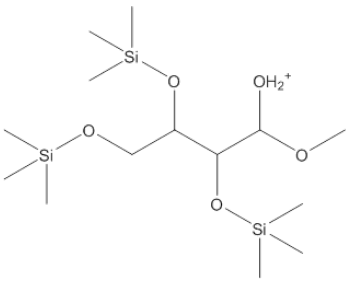
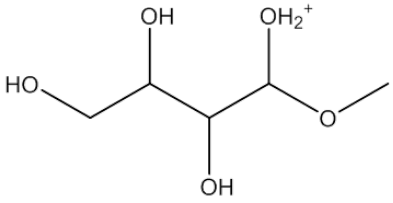
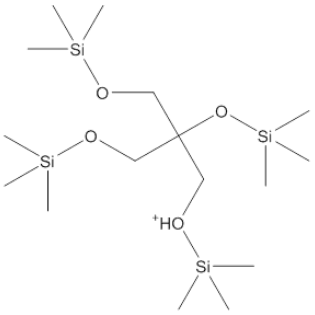
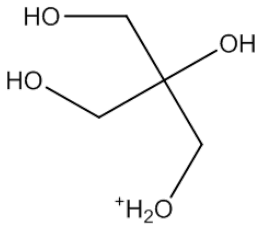
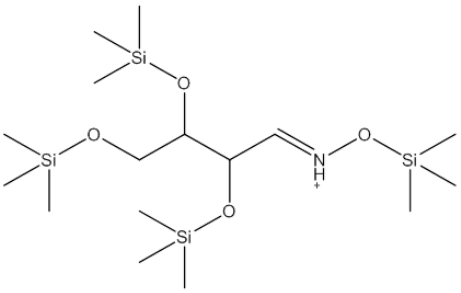
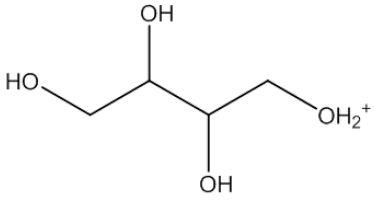
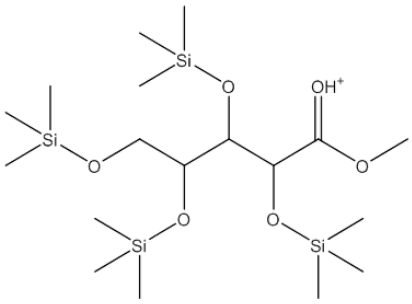
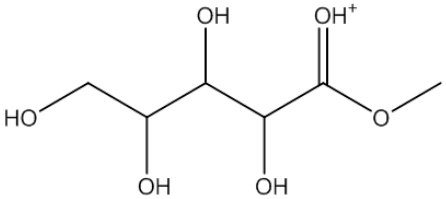
m/z	TMSO Structure	Original carbohydrate structure
323		
324		
325		
339		
350		
352		

Table R2.4. Cont.

m/z	TMSO Structure	Original carbohydrate structure
355		
369		
411		
424		
469		

Relevant ions in neutral monosaccharides spectra. Some relevant ions (*m/z* 309 and 324) could be derived from the fragmentation of all identified monosaccharides, although their abundances were different. Specifically, the abundance (up to 5–10%) of ion *m/z* 309 was high in arabinose and xylose while the highest abundance (up to 5–10%) of fragment *m/z* 324 was observed in glucose and galactose. Other ions could be formed only from some monosaccharide including fragment *m/z* 323 (abundance around 1%) which could originate from glucose, galactose and rhamnose.

Relevant ions in unidentified peaks spectra. With regard to these peaks the abundances of other ions were important. The most relevant ions for some of these peaks and their possible origin are indicated below.

U2 spectra. The abundance of ion *m/z* 195 (up to 10%), which could be formed from methyl-esterified GalA, was high.

U7 spectra. The abundance of ion *m/z* 424 (around 5%), which could be formed from apiose, was high.

U8 spectra. Ions *m/z* 234, 248, 250, 262, 350 and 352 showed high abundances (between 1 and 10%). These fragments could be derived from KDO.

U9 spectra. The ions with high abundances were *m/z* 197, 235, 339 and 469 (between 1 and 5%). Fragment *m/z* 197 could be derived from ferulic acid, present in artichoke (Lattanzio & van Sumere, 1987) and in sugar beet pulp pectin (Yapo & Koffi, 2008). While ion *m/z* 469 could be also formed from the fragmentation of methyl- and acetyl-esterified acidic sugars; ion *m/z* 235 could also be from rhamnose. Finally, fragment *m/z* 339 could also originate from all studied monosaccharides. U9 compound might be an acid molecule, maybe ferulic acid or an unknown acidic sugar. Probably, these two types of molecules coelute in the same U9 peak. This could explain the low accuracies and worse performances obtained for the classification of U9 spectra using simple models.

U11 spectra. The ion *m/z* 297 (abundance around 5%) could be formed from the fragmentation of methyl-esterified acidic sugars.

U12 spectra. Ions *m/z* 251, 325 and 355 were more abundant (abundances 1–5%). Ion *m/z* 251 could originate from rhamnose and *m/z* 325 and 355 from acidic sugars (acetylated or not).

U9 and U11 spectra. Important ions for both peaks were *m/z* 107 and 121 (abundances 0.1–0.5%). These two fragments could be derived from methyl-esterified acidic sugars, and ferulic/coumaric acid fragments.

U9 and U12 spectra. The abundance of ions *m/z* 209 (abundances between 0.1 and 0.5%) was higher in unidentified peaks 9 and 12 than in the rest of the peaks. This fragment could be derived from acidic sugar derivatives (methyl-esterified or not).

Relevant ions in several unidentified peaks spectra. Fragment *m/z* 133 (abundance 1–10%), also formed from rhamnose was high in U1, U2 and U11 peaks. Ion *m/z* 193 (abundance 1–5%) was high in U2, U5 and U12 peaks and could also originate from acidic sugars (methyl-esterified). Ion *m/z* 151 (abundance 1–3%), formed from acidic sugars, was high in U2, U7 and U9 peaks. Moreover, ion *m/z* 369 (high in U5 and U12; abundance 1–3%) could be formed from methyl-esterified acidic sugars. At last, ion *m/z* 411 (abundance 1–3%) was higher in U5 and U7 than in the rest of the peaks and could originate from apiose and most identified sugars.

These results suggest the presence of apiose (probably U7 peak), KDO (or similar molecules, probably U8 peak) and methyl-esterified (probably U2 and U11 peaks) or acetylated GalA or other acidic sugars (probably U1, U12 and other unidentified peaks) might be present in all pectin samples extracted by different treatments. It should be noted that the amounts of these unidentified monomers were higher in pectin extracted using enzymes (Table R2.3) and may be related with specific biological properties. This pectin also showed high amounts of known pectic sugars. Moreover, phenolic compounds like ferulic acid might be present in artichoke pectin, especially in pectin extracted with nitric acid (probably U9 peak), and might confer a potential antioxidant activity.

The ability to determine structural similarities and differences among pectins extracted by conventional and US-assisted methods as well as identified and unidentified compounds present in their structure could be of great importance to evaluate their functional activity. Given the great variability in the composition of the pectins according to the plant source and extraction methods, tools to classify these samples according to their structure may lead to a better understanding of their chemical

properties. In this regard, further research is needed to establish structure-function relationships.

CONCLUSIONS

Different pectin extraction methods have been studied. In the extraction time studied, 6 h, US combined with Celluclast produced the highest yields compared to US alone; however, GalA content was lower than 65%, showing that US treatments produced the highest extraction of other polysaccharides. With respect to molecule structure, pectin obtained with US + E and with Celluclast showed similar linearity pectin backbone and extent of branching of RG-I. Acid-extracted pectins presented a significantly higher GalA content. Taking into account these results a sequential combination of enzyme and US could be used to optimize the obtainment of pectin in short times, with high yields, low co-extraction of other polysaccharides, and without important molecular modification, except a lower GalA content.

To confirm the structural differences among extracted pectins, an ANN has been trained on pectin composition parameters, leading to high prediction rates on new samples. This model could be applied on pectin extracted from different sources and samples from unknown origin to determine similar hydrolysis patterns. In addition, GC-MS spectra modelling of both identified and unidentified sugars detected in pectin hydrolysates found highly reproducible classification patterns. The study of relevant *m/z* ions suggest that several unidentified sugars may correspond to acidic derivatives of acetyl- and methyl-esterified GalA.

Chapter II:

Pectic oligosaccharide production and characterization

Article III

6.2. Chapter II: Pectic oligosaccharide production and characterization

6.2.1. Article III: GC–MS characterization of novel artichoke (Cynara scolymus) pectic- oligosaccharides mixtures by the application of machine learning algorithms and competitive fragmentation modelling

Carbohydrate Polymers 205 (2019) 513–523

Contents lists available at ScienceDirect

Carbohydrate Polymers

journal homepage: www.elsevier.com/locate/carbpol

GC–MS characterization of novel artichoke (*Cynara scolymus*) pectic- oligosaccharides mixtures by the application of machine learning algorithms and competitive fragmentation modelling

Carlos Sabater, Agustín Olano, Nieves Corzo, Antonia Montilla

Instituto de Investigación en Ciencias de la Alimentación CIAL, (CSIC-UAM) CEI (UAM + CSIC), C/Nicolás Cabrera, 9, E-28049, Madrid, Spain

Link: https://www.sciencedirect.com/science/article/pii/S0144861718312438

A B S T R A C T

Novel artichoke pectic-oligosaccharides (POS) mixtures have been obtained by enzymatic hydrolysis using four commercial enzyme preparations: Glucanex®200G, Pentopan®Mono-BG, Pectinex®Ultra-Olio and Cellulase from *Aspergillus niger*. Analysis by HPAEC-PAD showed that Cellulase from *A. niger* produced the greatest

amount of POS (310.6 mg g^{-1} pectin), while the lowest amount was produced by Pentopan®Mono-BG (45.7 mg g^{-1} pectin). To determine structural differences depending on the origin of the enzyme, GC–MS spectra of di- and trisaccharides have been studied employing three machine learning algorithms: multilayer perceptron, random forest and boosted logistic regression. Machine learning models allowed characteristic *m/z* ions patterns to be established for each enzyme based on their GC–MS spectra with high prediction rates (above 95% on the test set). Possible chemical structures were given for some *m/z* ions having a decisive influence on these classifications. Finally, it was observed that several ions could be formed from specific POS structures.

Keywords: Artichoke pectin, Enzymatic hydrolysis, Pectic-oligosaccharides, Neural network, *In silico* fragmentation

INTRODUCTION

Pectin, one of the most structurally complex families of polysaccharides in nature, is mainly composed of linear chains of α-1,4-$_D$-galacturonic acid (GalA) called homogalacturonan (HG), which comprise approximately 70% of total pectin. The other two domains present ramified chains as rhamnogalacturonan type I (RG-I), a polymer with alternate sequences of GalA and α-(1, 2) linked $_L$-rhamnosyl residues which may be substituted at *O*-4 with linear or branched oligosaccharides, and rhamnogalacturonan II (RG-II), a most complex structure consisting of 12 different types of sugars in over 20 different linkages (Gullón et al., 2013; Mohnen, 2008). Pectins present interesting properties which are directly influenced by its structural characteristics (monomeric composition, presence and distribution of side chains, degree of methyl- and acetyl-esterification, molar mass, and charge distribution along their backbone) (Herbstreith & Fox, 2018; Maric et al., 2018).

The most commonly used pectins in the food industry are extracted from citrus peel, apple and sugar-beet by-products. However, the recent optimization of a procedure of

pectin extraction from artichoke by-products could provide a new source with interesting techno-functional properties (Sabater et al., 2018a).

By the partial depolymerization of pectins through chemical or enzymatic methods pectic-oligosaccharides (POS) can be obtained and it is known that POS are one of the most promising candidates to be recognized as prebiotics (Gullón et al., 2013). Enzymatic hydrolysis of pectins shows several advantages, such as high regio- and stereo-selectivity, obtaining structurally different POS mixtures (Babbar et al., 2016a; da Moura et al., 2015). Most food-grade POS studied are produced by commercial pectinases with three different enzymatic activities polygalacturonase (PG), pectin lyase (PL) and pectin esterase (PE) involving reactions of hydrolysis, β-elimination and removal of methyl or acetyl groups, respectively. Commercial pectinases are usually produced by *Aspergillus* spp. (Combo et al., 2012).

POS mixtures can be formed of various types of oligomers and these structures are often characterized using chromatographic techniques such as HPAEC-PAD (Babbar et al., 2017; Combo et al., 2013; Gómez et al., 2013) and mass spectrometry (MS) (Ognyanov et al., 2016; Leijdekkers et al., 2015). MS allows structural patterns of complex molecules to be determined and has become the analytical method of choice in metabolomics research. However, it generates large amounts of intricate data that needs to be interpreted to extract chemical information and ensure that they are of valuable (Yi et al., 2016; Boccard & Rudaz, 2014). The identification of unknown compounds is the main bottleneck, so computational tools assisting structure elucidation and *de novo* identification of small molecules have been developed including *in silico* fragmentation methods (Allard et al., 2016; Ruttkies et al., 2016). These methods help in the deduction of possible structures of metabolites in spectra interpretation. *Competitive fragmentation modelling* (CFM) has been proposed as an *in silico* fragmentation method suitable for common MS techniques such as GC-EI-MS and significantly outperforms existing state-of-the-art computational methods (Allen et al., 2016; Allen et al., 2015). It has been validated and tested on the NIST database, and it could be used as an alternative when no reference standard is available for measurement.

Other computational tools are focused on data modelling to find reproducible patterns and discover valuable information on biological events and chemical/structural

properties (Yi et al., 2016; Käll et al., 2007). A number of machine learning classification methods have been applied in the analysis of MS spectra. Supervized and semi-supervized classification methods support *a priori* known data structures to train patterns and rules to predict new data. When the relationship between MS data and chemical structures is complex, simple classifiers may not be efficient. Today there exists a multitude of algorithms to determine classification patterns among samples, including support vector machines (SVM) and random forests (RF), which can also be applied to the field of metabolomics (Lin et al., 2014; Uarrota et al., 2014) and also food science (Sabater et al., 2018b; Lim et al., 2017). Tree-based models have been used for proteomic mass spectra classification (Geurts et al., 2005) and recent works reported high prediction rates based on MALDI-TOF data (Rossel & Martínez Arbizu, 2018). Artificial neural networks (ANN) and boosted models are other machine learning algorithms used to classify spectral data of small molecules (Gosav et al., 2011), which recognize chemical substructures from MS data (Varmuza et al., 2003), feature selection in MS-based proteomic profiling (Gertheiss & Tutz, 2009) or protein biomarker discovery (Yasui et al., 2003). It has been reported that machine learning has greatly improved performance relative to the bond-breaking approaches and even CFM (Schymanski et al., 2017).

Therefore, the aim of this study was characterized by GC–MS novel artichoke POS obtained by enzymatic hydrolysis using commercial enzyme preparations with different main activities. In order to accomplish this, structural characteristics of di- and tri-POS were determined by mass spectral mining using three machine learning algorithms: multilayer perceptron (MLP), random forest (RF) and boosted logistic regression (BLR).

MATERIALS AND METHODS

Standards and reagents

Analytical reference substances such as sucrose, $_{D}$-arabinose, $_{D}$-xylose, $_{L}$-rhamnose, $_{D}$-galactose, $_{D}$-mannose, $_{D}$-glucose, galacturonic acid (GalA), digalacturonic acid (Di-GalA), kestose, nystose and β-phenyl glucoside were purchased from Sigma Aldrich (Steinheim, Germany). Trigalacturonic acid (Tri-GalA) was from Carbosynth (Compton, UK). Four commercial enzyme preparations were studied (Table R3.1). Cellulase from *Aspergillus niger* was acquired from Sigma Aldrich (Steinheim, Germany). The rest of the enzyme preparations were a generous gift from Novozymes (Bagsvaerd, Denmark). Artichoke pectin was previously extracted in our laboratory using a commercial cellulase, Celluclast®1.5 L (artichoke by-product powder concentration 6.5%, enzyme dose 10.1 U g^{-1}, extraction time 48 h). This pectin has GalA content of 69.5%, degree of methyl-esterification of 19.5% and molecular mass (M_w) ranging from 4.8 to 660 kDa (Sabater et al., 2018a).

Enzyme characterization

Enzyme characterization assays were carried out following the method described by Martínez et al. (2009a,b) with modifications. A solution of 2% (w/v) of polygalacturonic acid, chosen as HG standard (Sigma, purity > 85%, GalA content greater than 96%), was dissolved in 0.05 M acetate buffer (pH 5.0). Pectinase activity was measured by quantifying the amount of reducing sugar groups liberated after incubation of polygalacturonic acid solutions with 5 mg mL^{-1} or 5 μL mL^{-1} of enzyme at 50 °C for 5 min, using the method of DNS and GalA as standard. One unit of pectinase activity was defined as the amount of enzyme required to release 1 μmol of GalA per min at 50 °C.

Formation of pectic-oligosaccharides (POS)

Pectic-oligosaccharides (POS) were obtained by enzymatic hydrolysis of 2% (w/v) artichoke pectin solutions dissolved in 0.05 M acetate buffer (pH 5.0) using 0.54–6.75 U mL^{-1} of enzyme (Table R3.1) following the method of Gómez et al. (2016) with some modifications. Enzymatic hydrolysis was performed in a final volume of 1 mL incubated in an orbital shaker at 50 °C and 750 rpm. Aliquots were withdrawn from the reaction mixture at the different times (0.5, 1 and 4 h) and immediately immersed in boiling water for 5 min to inactivate the enzyme. Samples were stored at −18 °C for subsequent analysis. Enzymatic reactions were carried out in duplicate and analyses were performed twice for each enzymatic treatment. In addition, four hydrolysis replicates were prepared for reactions incubated at 4 h for their GC–MS characterization.

Table R3.1. Determination of commercial enzyme preparation activities (U g^{-1} and U mL^{-1}) using polygalacturonic acid as substrate.

Enzyme preparation	**Microorganism**	**Declared activity**	**Hydrolase activity (U g^{-1})**	**Enzyme dose** (U mL^{-1})**
Pectinex®Ultra-Olio	*Aspergillus aculeatus/ Aspergillus niger*	Pectin-lyase	202.6 ± 7.4[a,b]*	6.75
Glucanex®200G	*Trichoderma harzianum*	Glucanase	189.2 ± 28.5[b]	0.63
Pentopan®Mono-BG	*Thermomyces lanuginosus*	1,4-endoxylanase	161.5 ± 3.6[b]	0.54
Cellulase from *A. niger*	*Aspergillus niger*	Cellulase	263.7 ± 18.5[a]	0.88

*Enzyme preparation in liquid form: U mL^{-1}.

** Enzyme dose used for POS production

[a,b] Statistically significant differences between enzymes

Analytical techniques

Monosaccharides were quantified by GC-FID as trimethyl silylated oximes (TMSO) (Sabater et al., 2018a). In addition, samples containing di- and tri-POS formed were analysed by GC–MS on an Agilent Technologies 7890 A gas chromatograph coupled to a 5975CMSD quadrupole mass detector (Agilent Technologies) to characterise low M_w POS obtained. Separations were carried out using a fused silica capillary column HP-5MS (5% phenyl methylsilicone, 30 m × 0.25 mm × 0.25 μm thickness; J&W Scientific, Folsom, CA, USA). Helium was used as carrier gas at a flow rate of 0.8 mL min^{-1}. Injector temperature was 200 °C. The oven temperature was initially 200 °C and held for 5 min, then increased at a rate of 3 °C min^{-1} to 250 °C and held for 1 min, then increased at a rate of 10 °C min^{-1} to 320 °C and held for 70 min. Injections were made in the split mode (1:5). The mass spectrometer was operated in electrospray ionization mode at 70 eV. Mass spectra were acquired using Agilent ChemStation MSD software. Internal standard (β-phenyl glucoside) was added to the samples. Identification of trimethylsilyl oxime derivatives of carbohydrates was carried out by comparison of their relative retention times and mass spectra with those of standard compounds (GalA, Di-GalA and Tri-GalA).

Average M_w distribution of POS formed were determined by HPSEC-ELSD following the methods described by Sabater et al. (2018a). For HPSEC-ELSD analysis, samples (0.65 mg mL^{-1}) were dissolved in water filtered using a 0.45 μm syringe filter (Symta, Spain) and analysed in an Agilent Technologies 1220 Infinity LC System (Böblingen, Germany) The separation of carbohydrates was carried out on a TSK-GEL G5000PW_{XL} column (300 mm x 7.8 mm, 10 μm particle size) and TSK-GEL G2500PW_{XL} column (300 mm x 7.8 mm, 6 μm particle size) in combination with a TSK-GEL PW_{XL} guard column (40 mm x 6 mm, 12 μm particle size) (Tosoh Bioscience, Montgomeryville, PA, USA) using 0.01 M NH_4Ac, as mobile phase and elution in isocratic mode at a flow rate of 0.5 mL min^{-1} for 50 min. M_w of carbohydrates was calculated by the external calibration method using solutions of commercial pullulan standards (M_w 0.342–788 kDa) (Fluka Analytical) in the range 10–2250 mg L^{-1}.

POS obtained with different commercial enzyme preparations were quantified by HPAEC-PAD using a DIONEX ICS2500 system (Dionex Corp., Sunnyvale, CA, USA) incorporating a GP50 gradient pump and an ED50 electrochemical detector using a gold and Ag/AgCl as working and reference electrodes, respectively. Separations were carried out at room temperature in a CarboPac PA-1 column (2 × 450 mm) and a CarboPac PA 1 guard column (4 × 50 mm) as a flow rate of 1 mL min^{-1}. The mobile phases were (A) 0.1 M NaOH and (B) 1 M NaOAc in 0.1 M NaOH). The elution profile was as follows: 0-15 min, 0–5% B; 15–60 min, 5–70% B; 60–65 min, 70–100% B; 65–70 min, 100% B; 70–70.1 min, 100-0% B; and finally column re-equilibration by 0% B from 70.1 to 85 min. Pectin neutral monosaccharides (arabinose, xylose, rhamnose, galactose), neutral di- tri- and tetrasaccharides (sucrose, kestose and nystose) as well as GalA and its derivatives (Di-GalA and Tri-GalA) standards were used for identification. Acquisition and processing of data were achieved with Chromeleon 6.7 software (Dionex Corp. CA, USA).

Data analysis

ANOVA tests and Tukey´s test for $p < 0.05$ were applied to all data generated. Structural characteristics of POS obtained with Pentopan®Mono-BG, Glucanex®200 G, Cellulase from *A. niger* and Pectinex®Ultra-Olio were determined by GCeMS spectral mining. In order to find reproducible patterns which could be applied on new samples, three machine learning models were evaluated in a supervized classification task: multilayer perceptron (MLP), random forest (RF) and boosted logistic regression (BLR). To ensure these algorithms were trained with valuable information, *m/z* fragments whose abundances were statistically different among groups ($p < 0.05$) corresponding to known POS ruptures were selected.

Signal processing

To increase model performance, discrete wavelet transform (DWT) was applied to GC–MS data (Li et al., 2007; Xia et al., 2007). In this study, a 16 level DWT was

computed (indicating the depth of the decomposition) with a “la8” decomposition filter (Daubechies orthonormal compactly supported wavelet of length = 8). In the threshold step, DWT coefficients under percentile 15% were removed to denoise the signal. Once the signal was reconstructed, all variables were scaled and centered.

Classification of MS spectra

462 mass spectra of known pectic sugars and unknown POS released during enzymatic hydrolysis were classified using MLP, RF and BLR.

MLP is the most common kind of artificial neural network (ANN), a family of broadly used models that allow modeling complex and highly non-linear processes. MLP is formed by an input layer (i.e. preprocessed GC–MS spectra), an output layer (i.e. enzyme used to obtain POS) and several neurons or nodes organized in hidden layers, where each neuron in a layer is connected with each neuron in the next layer through a weighted connection. In this case, an MLP was built with 1 hidden layer consisting in 25 neurons. The activation function (a transformation applied to the input signal to determine whether the information that the neuron is receiving is relevant or not) was logistic. In RF, a multitude of decision trees are constructed, outputting the different classes. Each node is split using the best among a subset of predictors randomly chosen. In this case, RF model was built with 500 trees and 50 variables tried at each split.

On the other hand, BLR is considered as an ensemble method that uses a weighted average of predictions of individual classifiers (Geurts et al., 2005). The iterations specify the maximal number of trees to be fitted. In this work, a BLR consisting in 11 iterations was chosen.

All the models were trained with 70% of the data, 10-fold cross-validated and then tested with 30% of data from each group (corresponding to new samples). Variable importance analysis was carried out to determine the most influential *m/z* ions in each model. In MLP, influent *m/z* ions were determined by the sum of the product of raw input-hidden and hidden-output connection weights while a permutation of the out of-

bag-error was chosen for RF (an estimation of the classification error as trees are added to the forest). For BLR, importance coefficients were determined by calculating the area under the ROC (Receiver Operating Characteristic) curve.

All statistical analyses were computed on R v3.5.0. DWT was performed using wavelets package (Aldrich, 2013). MLP was computed using the RSNNS package (Bergmeir & Benitez, 2012) and RF classification was performed with Random Forest package (Liaw & Wiener, 2002). For BLR, caTools package was employed (Tuszynski, 2014).

In silico fragmentation

After determining the most important *m/z* ions in di- and tri-POS classification, chemical structures of some fragments have been proposed employing the competitive fragmentation modeling source code (CFM-ID) developed by Allen et al. (2016). GC-EI-MS fragmentation patterns of POS structures described in the bibliography (Atmodjo et al., 2013) were determined and compared to those of enzymatically obtained POS.

RESULTS AND DISCUSSION

The polygalacturonase activity of four commercial enzyme preparations using polygalacturonic acid as a substrate was studied. As can be seen in Table R3.1, Pectinex Olio, Cellulase from *A. niger*, Glucanex and Pentopan showed high activities (202.6 U mL^{-1}, 263.7, 189.2 and 161.5 U g^{-1}, respectively. Moreover, these enzymes have different declared enzymatic activities including cellulase, pectin-lyase, exo-β-$_D$-galactofuranosidase and 1, 4-endoxylanase, respectively and, therefore, different hydrolysis patterns could be expected. The complementary activities could be of great importance, so Cellulase from *A. niger* presented higher polygalacturonase activity than cellulase activity using polygalacturonic acid and carboxymethyl cellulose as substrates (Sabater et al., 2018a).

Pectic-oligosaccharides (POS) obtainment from artichoke pectin hydrolysis

Enzymatic obtainment of POS derived from artichoke pectin using the four tested enzymes was studied. In Table R3.1 the dose of each enzyme used to hydrolyse artichoke pectin is shown. Monosaccharides released during enzymatic hydrolysis were quantified by GC-FID. Figure R3.1 shows the amounts found of each compound, GalA, neutral monosaccharides derived from pectins, such as xylose, arabinose, rhamnose and galactose as well as other unidentified monosaccharides present in artichoke hydrolysates. Pectinex Olio released significantly higher amounts of GalA (14.8 mg 100 mg^{-1} pectin) and neutral sugars (11.5 mg 100 mg^{-1} pectin), as expected, due to the high amount of added enzyme (6.75 U mL^{-1}), to attempt obtaining a large amount of low M_w POS. This elevated GalA release may also be due to their declared activity, pectin-lyase. In the other enzymatic preparations, although the enzyme dose was similar at 0.54, 0.63 and 0.88 U mL^{-1} of reaction mixture, the hydrolysis patterns of artichoke pectin were very different, probably because of their different main activities. Therefore, Cellulase from *A. niger* released a significantly major amount of GalA and Glucanex released mainly arabinose (in amounts significantly higher than Cellulase from *A. niger* and Pentopan but still lower than the ones obtained with Pectinex Olio). Lower hydrolysis rates were expected for Glucanex considering its main activity (glucanase) which may not produce significant ruptures in the pectin backbone. Results obtained with Cellulase from *A. niger* indicate significant pectinase secondary activity, even higher than its declared activity. On the contrary, the hydrolytic activity of Pentopan (α-1,4-endoxylanase) was very low showing a significantly lower monosaccharide release, probably due to a lower presence of xylose in artichoke pectin ramified chains. Interestingly, Cellulase from *A. niger* and Pectinex Olio also released 0.2-0.5 mg 100 mg^{-1} pectin of unidentified monosaccharides probably due to other secondary enzymatic activities. Neutral monosaccharides released from artichoke pectin hydrolysis, mainly arabinose and galactose, depict the relevance of a side chain of RG-I structures, possibly arabinans, galactans and arabinogalactans.

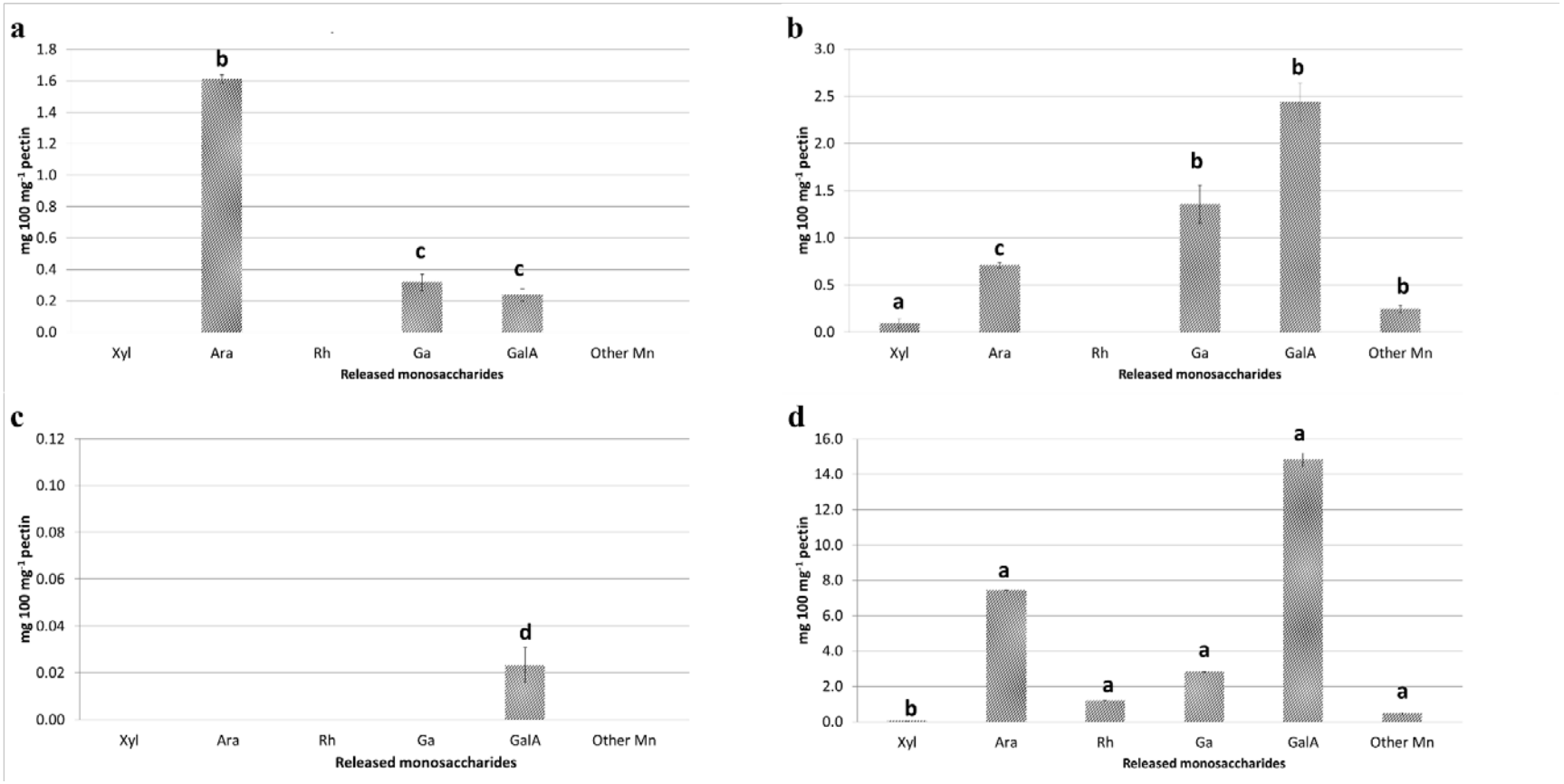


Figure R3.1. Monosaccharides (mg 100 mg^{-1} pectin) found in enzymatic hydrolysates of artichoke pectin after 4 h of reaction using: **a)** Glucanex® 200G (0.63 U mL^{-1}); **b)** Cellulase from *Aspergilus niger* (0.88 U mL^{-1}); **c)** Pentopan® Mono-BG (0.54 U mL^{-1}); **d)** Pectinex® Ultra-Olio (6.75 U mL^{-1}). Xyl: xylose, Ara: arabinose, Rh: rhamnose, Ga: galactose, GalA: galacturonic acid, Other Mn: unidentified monosaccharides. ^{a,b,c,d} Statistically significant differences between enzymes.

As previously reported in Sabater et al. (2018a), artichoke pectin contains a higher percentage of neutral monosaccharides compared to pectin extracted from apple and citrus pectin (Bonnin et al. 2014). Considering these differences, artichoke POS would be expected to show different structures to those derived from other pectin sources.

M_w distribution of enzymatic hydrolysates was determined by HPSEC-ELSD. Pectin from artichokes showed three main fragments with M_w of 660, 105 and 4.8 kDa, as has been previously reported (Sabater et al., 2018a) and after enzymatic treatments, differences in the chromatographic profiles were observed depending on the enzyme selected (Figure R3.2). As can be seen, Pectinex Olio and Cellulase from *A. niger* (Figure R3.2b) showed the most different patterns of POS average M_w distribution compared to the original product (Figure R3.2a), producing the release of a wider variety of fragments and higher amounts of POS with M_w between 50 and 3 kDa (fractions with $M_w < 3$ kDa were not quantified due to coelution with other compounds present in enzymatic preparations). In contrast, chromatographic profiles of Pentopan hydrolysates showed hardly any modification with respect to artichoke pectin while Glucanex produced an important reduction de M_w producing modified pectins ($M_w \sim 50$ kDa) (Figure R3.2a and c). In general, average M_w distribution of POS produced with these preparations varied from 78 to 3.5 kDa (Figure R3.3). Enzymatic hydrolysis with Cellulase from *A. niger* produced, after 1 h of reaction, POS of M_w around 14 kDa (60.3% of fragments determined) and the M_w decreased slightly after 4 h up to 11 kDa, while Pectinex Olio formed high amounts of POS with lower M_w (5.7–6.0 kDa, 77–80% of fragments determined). As has been stated, no changes were observed in the HPSEC-ELSD profiles of the other of hydrolysates (Figure R3.2c). These results highlight the importance of enzymatic activity in POS structural characteristics and several studies reported the influence of enzyme and reaction conditions over POS formation (Babbar et al., 2016a; Combo et al., 2012).

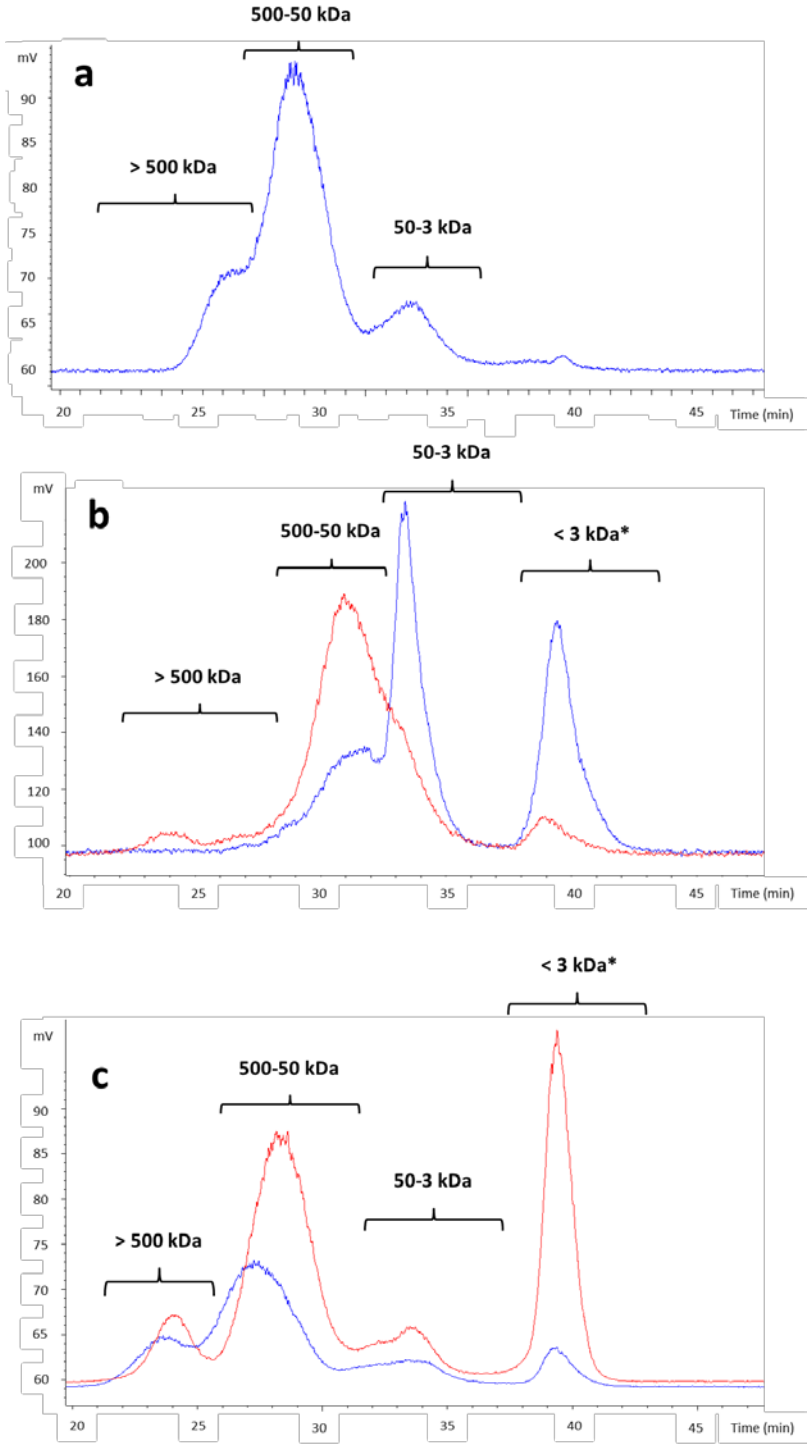


Figure R3.2. HPSEC-ELSD profiles and molecular weight ranges of a) artichoke pectin; and hydrolysates obtained from artichoke pectin after 4 h of reaction using b) Cellulase from *A. niger* (red) and Pectinex®Ultra-Olio (blue), c) Pentopan®Mono-BG (blue) and Glucanex®200G (red). *Compounds not quantified due to coelution with other present in enzymatic preparations.

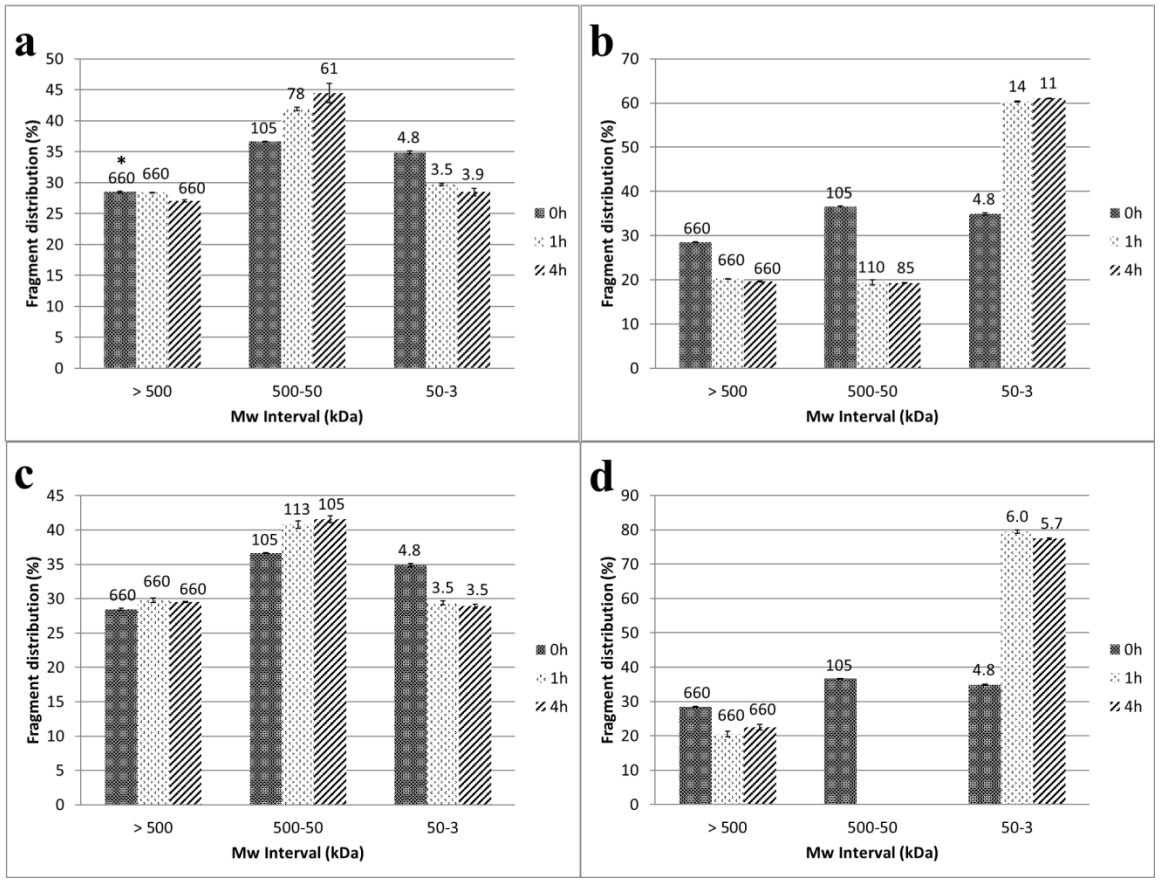


Figure R3.3. Molecular weight distribution of fragments with M_w > 3 kDa determined by HPSEC-ELSD of different hydrolysates of artichoke pectin after 1 and 4 h of reaction using: **a)** Glucanex® 200G (0.63 U mL^{-1}); **b)** Cellulase from *Aspergilus niger* (0.88 U mL^{-1}); **c)** Pentopan® Mono-BG (0.54 U mL^{-1}); **d)** Pectinex® Ultra-Olio (6.75 UmL^{-1}). *Average M_w of the fragments within M_w interval.

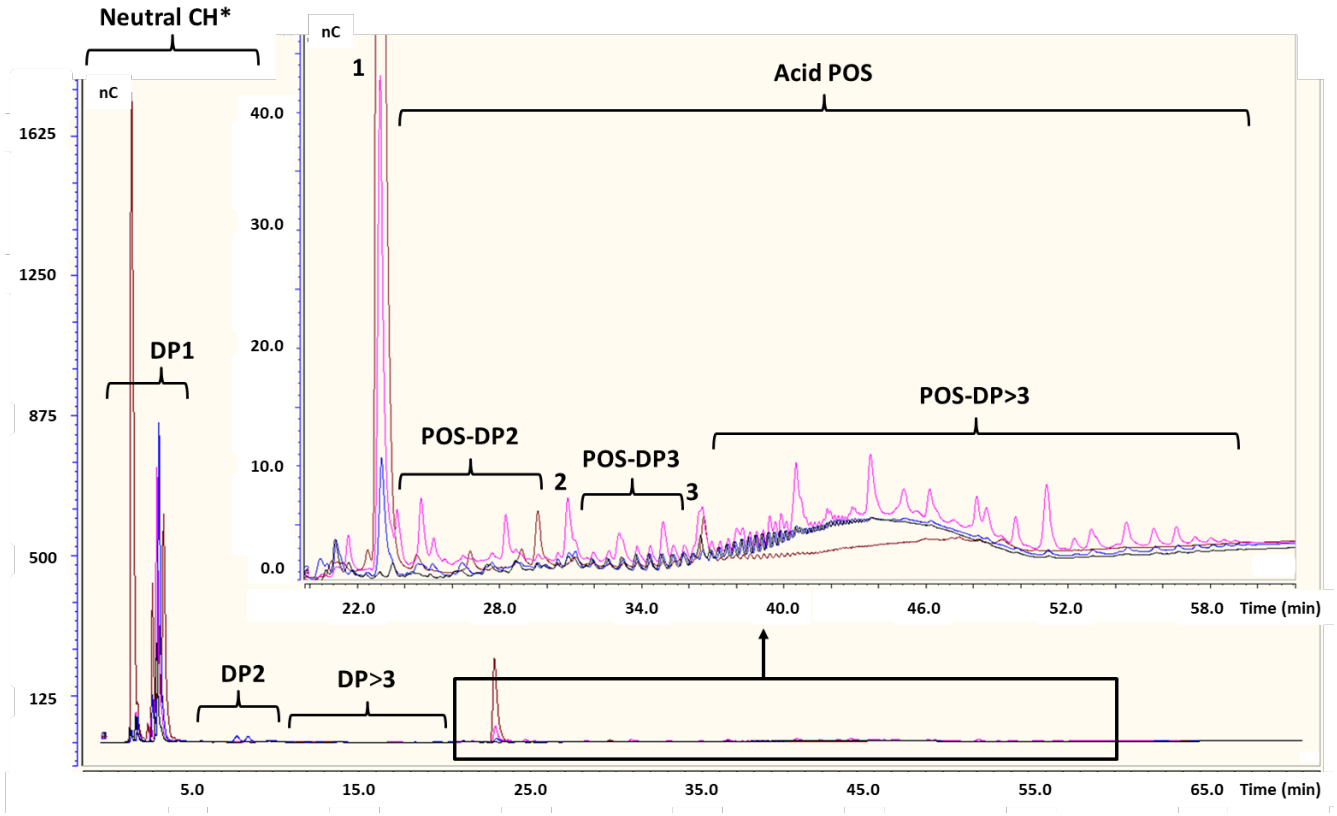


Figure R3.4. HPAEC-PAD profiles of carbohydrates found in artichoke pectin hydrolysates using commercial enzyme preparation Cellulase from *A. niger* (pink), Pectinex Olio (brown), Glucanex (blue) and Pentopan (black) (50 ºC, pH =5.0, 4h, 0.54-6.75 U mL^{-1} enzyme). CH: carbohydrates, POS: pectic-oligosaccharides. Peaks: DP1, DP2 and DP3 neutral mono-, di- and trisaccharides DP > 3, neutral oligosaccharides; (1) Galacturonic acid; (2) Digalacturonic acid; (3) Trigalacturonic acid; POS-DP2, acid disaccharides; POS-DP3, acid trisaccharides; POS-DP>3, others acid oligosaccharides. (*) Compounds not quantified.

Table R3.2. POS formation (HPAEC-PAD) and galacturonic acid (GalA) release (mg g^{-1} pectin) during enzymatic hydrolysis (0.5, 1 and 4h) of artichoke pectin using the four studied enzymes. Di-GalA: digalacturonic acid, Tri-GalA: trigalacturonic acid, POS: pectic-oligosaccharides.

Hydrolysis	**Time (h)**	**GalA (peak 1)**	**Acid POS (mg g^{-1} pectin)**					
			POS-DP2	**Di-GalA (peak 2)**	**POS-DP3**	**Tri-GalA (peak 3)**	**POS-DP > 3**	**Total POS***
Pentopan®Mono-BG	0.5	3.0 ± 0.2 ^{d}	12.2 ± 1.0 ^{d}	7.8 ± 0.6 ^{b}	6.4 ± 0.2 $^{c, d, e}$	7.0 ± 0.5 $^{a, b}$	-	33.4 ± 2.8 ^{e}
	1	6.5 ± 0.6 ^{d}	12.6 ± 0.9 ^{d}	8.0 ± 0.6 ^{b}	7.7 ± 0.5 $^{c, d}$	7.4 ± 0.2 $^{a, b}$	-	35.7 ± 2.5 ^{e}
	4	10.0 ± 0.7 ^{d}	12.8 ± 1.1 ^{d}	8.2 ± 0.6 ^{b}	7.3 ± 0.2 $^{c, d}$	7.4 ± 0.5 $^{a, b}$	-	45.7 ± 3.9 ^{e}
Glucanex®200G	0.5	6.5 ± 0.6 ^{d}	17.9 ± 1.3 $^{b, c, d}$	7.3 ± 0.5 ^{b}	0.1 ± 0.0 ^{e}	6.2 ± 0.2 $^{b, c}$	9.6 ± 0.8 ^{c}	41.1 ± 2.9 ^{e}
	1	11.0 ± 0.8 ^{d}	20.2 ± 1.4 $^{b, c}$	7.3 ± 0.5 ^{b}	4.6 ± 0.1 $^{d, e}$	7.7 ± 0.5 $^{a, b}$	22.1 ± 1.6 ^{c}	61.9 ± 5.3 $^{d, e}$
	4	11.4 ± 1.0 ^{d}	23.3 ± 1.6 $^{a, b}$	7.7 ± 0.5 ^{b}	6.9 ± 0.5 $^{c, d, e}$	7.1 ± 0.2 $^{a, b}$	18.7 ± 1.6 ^{c}	63.7 ± 4.5 $^{d, e}$
Cellulase from *A. niger*	0.5	15.0 ± 1.1 ^{d}	17.9 ± 1.3 $^{b, c, d}$	7.1 ± 0.5 ^{b}	11.7 ± 0.3 ^{c}	7.2 ± 0.5 $^{a, b}$	184.5 ± 13.0 ^{b}	228.4 ± 19.4 ^{b}
	1	22.6 ± 1.9 ^{d}	20.2 ± 1.4 $^{b, c}$	7.3 ± 0.5 ^{b}	9.1 ± 0.6 $^{c, d}$	7.6 ± 0.2 $^{a, b}$	196.2 ± 16.6 $^{a, b}$	240.4 ± 17.0 ^{b}
	4	43.0 ± 3.0 ^{d}	23.3 ± 2.0 $^{a, b}$	7.4 ± 0.5 ^{b}	47.3 ± 1.3 ^{b}	7.4 ± 0.5 $^{a, b}$	225.2 ± 15.9 ^{a}	310.6 ± 26.4 ^{a}
Pectinex® Ultra-Olio	0.5	247.3 ± 21.0 ^{c}	16.0 ± 1.1 $^{c, d}$	5.8 ± 0.4 ^{b}	53.5 ± 3.8 $^{a, b}$	8.4 ± 0.2 ^{a}	13.4 ± 1.1 ^{c}	97.1 ± 6.9 $^{c, d}$
	1	382.1 ± 27.0 ^{a}	22.1 ± 1.9 ^{b}	13.8 ± 1.0 ^{a}	55.6 ± 1.6 ^{a}	8.3 ± 0.6 ^{a}	15.2 ± 1.1 ^{c}	115.0 ± 9.8 ^{c}
	4	303.8 ± 25.8 ^{b}	29.0 ± 2.1 ^{a}	16.1 ± 1.1 ^{a}	55.2 ± 3.9 ^{a}	5.0 ± 0.1 ^{c}	23.2 ± 1.6 ^{c}	128.5 ± 9.1 ^{c}

***Total POS**: Σ POS-DP2, Di-GalA, POS-DP3, Tri-GalA and POS-DP > 3. ^{a,b,c,d,e} Statistically significant differences between groups.

Differences in hydrolysis patterns were also observed in the HPAEC-PAD analysis. Figure R3.4 shows a chromatographic profile of carbohydrates found in enzymatic hydrolysate of artichoke pectin after 4 h of reaction with enzyme preparations. Neutral sugars were eluted before acid sugars, GalA (peak 1), Di-GalA (peak 2), Tri-GalA (peak 3), and unknown acid POS with degree of polymerization (DP) of 2, 3 and > 3 were found. In comparison to retention times with commercial standards, these compounds could correspond to oligosaccharides with one/two molecules of GalA and one or more molecules of neutral sugars. Total acid POS formed with the four enzymes used were quantified (Table R3.2) and ranged from 33.4 to 310.6 mg g^{-1} pectin. Pectinex Olio released significantly higher amounts of GalA, Di-GalA and Tri-GalA. Interestingly, Cellulase from *A. niger* formed significant higher amounts of larger oligosaccharides. Maximum formation of POS was reached at 4 h of enzymatic hydrolysis, with a POS yield of 45.7, 63.7, 128.5 and 310.6 mg g^{-1} pectin for Pentopan, Glucanex, Pectinex Olio and Cellulase from *A. niger*, respectively.

GC–MS characterization of POS-DP2 and POS-DP3 obtainment from artichoke pectin

To delve into structural characteristics of novel artichoke POS mixtures released during enzymatic hydrolysis, a GC–MS study of these oligosaccharides was carried out. In Figure R3.5a CG-MS profile of hydrolysate of artichoke pectin using Cellulase from *A. niger* is shown. Di-GalA and Tri-GalA as well as unknown POS-DP2 and POS-DP 3 were detected. Similar GC–MS profiles have been obtained for the other enzymes used.

Interestingly, specific *m/z* ions 277, 321, 333 and 423, derived from β-cleavage fragmentation of uronic acids, as well as GalA characteristic ions *m/z* 332 and 540 were detected in Di-GalA MS spectra (Füzfai et al., 2004; Petersson, 1974). Taking into account the scarce information found in the bibliography about the GC–MS spectra of this type of compound a mass spectral study employing three machine learning algorithms (multilayer perceptron, MLP; random forest, RF; and boosted logistic regression, BLR) was carried out. Therefore, full mass spectra of all disaccharides and trisaccharides and unknown POS found in the enzymatic hydrolysates were classified.

The number of spectra included in this study is shown in Table R3.3. A total of 462 MS spectra were used for data analysis and classified according to the enzyme used, Glucanex-unknown POS (n = 104); Cellulase from *A. niger*-unknown POS (n = 116); Pentopan-unknown POS (n = 44); Pectinex Olio-unknown POS (n = 128) and different known pectic sugars (arabinose, rhamnose, xylose, galactose, GalA and Di- and Tri-GalA, n = 70).

For data analysis, 151 fragments in the range of *m/z* 61–546, which were statistically different among groups ($p < 0.05$) and might correspond to known POS ruptures, assessed by CFM-ID (Allen et al., 2016), were selected. Then, each MS spectra was decomposed and reconstructed using the DWT. Unknown POS were classified according to the enzyme used (Glucanex, Pentopan, Cellulase from *A. niger* and Pectinex Olio), and were also differentiated from known pectic sugars. These models were validated and tested on 30% of new samples. The training, 10-fold cross-validation and test rates were:

MLP: 100, 91.1 and 95.7%, respectively.

RF: 100, 97.8 and 100%, respectively.

BLR: 100, 98.1 and 100%, respectively.

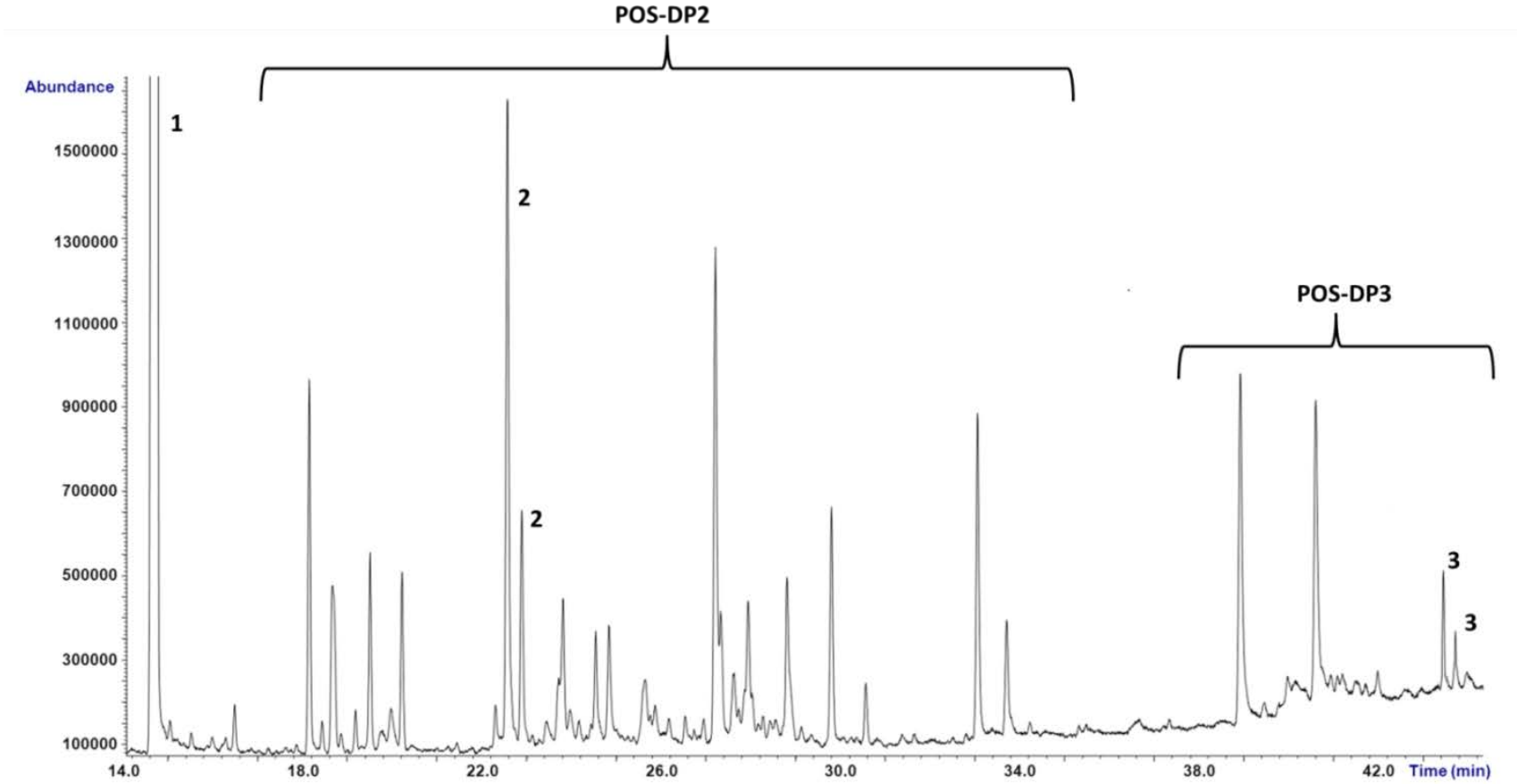


Figure R3.5. GC-MS profile of hydrolysate of artichoke pectin obtained by incubation with cellulase from *A. niger*. Peaks: (1) Internal standard, (2) Digalacturonic acid (Di GalA), (3) Trigalacturonic acid (Tri-GalA). POS-DP2: unknown pectic disaccharides, POS-DP3: unknown pectic trisaccharides.

Table R3.3. Number of GC-MS spectra (n = 462) of oligosaccharides obtained from GC-EI-MS analysis of hydrolysates from artichoke pectins included in machine learning study. GalA: galacturonic acid, Di-GalA: digalacturonic acid, Tri-GalA: trigalacturonic acid, POS: pectic-oligosaccharides.

Artichoke pectin + enzyme	**Commercial standards**	**Number of hydrolysates**	**Number of GC-MS spectra**					
			Known pectic monosaccharides	**Di-GalA**	**Unknown POS-DP2**	**Tri-GalA**	**Unknown POS-DP3**	**Total**
Glucanex®200G	-	4	-	8 (2 per hydrolysate)	60 (15 per hydrolysate)	8 (2 per hydrolysate)	44 (11 per hydrolysate)	16 known POS 104 unknown POS
Cellulase from *A. niger*	-	4	-	8 (2 per hydrolysate)	84 (21 per hydrolysate)	8 (2 per hydrolysate)	32 (8 per hydrolysate)	16 known POS 116 unknown POS
Pentopan®Mono-BG	-	4	-	8 (2 per hydrolysate)	24 (6 per hydrolysate)	8 (2 per hydrolysate)	20 (5 per hydrolysate)	16 known POS 44 unknown POS
Pectinex®Ultra-Olio	-	4	-	8 (2 per hydrolysate)	112 (28 per hydrolysate)	-	16 (4 per hydrolysate)	8 known POS 128 unknown POS
-	Standard	-	10[a]	2[b]	-	2[c]	-	14

[a] Known monosaccharides including galacturonic acid, arabinose, rhamnose, xylose and galactose. [b] Digalacturonic acid standards. [c] Trigalacturonic acid standards.

As can be observed, machine learning algorithms found a highly reproducible classification pattern that could be applied on new samples. DWT preprocessing allowed model prediction rates to increase up to 95–100%. When these models were tested on new samples, MLP showed false negative rates of 0.72, 1.45 and 2.17% for POS obtained with Glucanex, Pentopan and Pectinex Olio preparations, respectively. The false positive rates were 0.72% for known pectic sugars, also for POS obtained with cellulase from *A. niger*, Pentopan and Pectinex Olio were 0.72% and 1.60% for POS from Glucanex preparations. Although MLP showed high accuracy in its predictions (above 95%), RF and BLR correctly classified 100% of test samples. Therefore, RF and BLR showed 100% sensitivity, specificity and balanced accuracy values for each class. In contrast, MLP showed test sensitivity values between 85.7–96.7% for Pentopan, Pectinex Olio and Glucanex POS and 100% for Cellulase from *A. niger* POS and known sugars. MLP specificity values ranged from 98.2 to 99.2% for all the classes and balanced accuracy between sensitivity and specificity was 92.5, 95.7, 97.4, 99.5 and 99.6% for Pentopan, Pectinex Olio, Glucanex and Cellulase from *A. niger* POS and known sugars, respectively. Interestingly, MLP possessed higher specificity for POS classification, obtaining the lowest accuracy for Pentopan POS. However, all these values were above 90%. Higher accuracy for known sugars might be due to structural similarities among POS mixtures. The most influential *m/z* ions and their importance in each model were determined (Supplementary material Tables S3.1-S3.3).

In addition, overall accuracy and kappa values obtained with each model via their resampling distributions were compared (a comparative account is shown in Supplementary material Figure S3.1). The accuracy indicates the number of instances that were classified correctly, while kappa is a more robust measure that takes into account the possibility of a correct classification by chance. Both values were similar in the three models. BLR and RF showed mean accuracy and kappa values of 0.97-0.98, significantly higher than those of MLP (0.91 and 0.89, respectively). In all three models, these metrics were high, indicating high predictive power.

To deepen into POS structural differences established by these models, a total 50 *m/z* values of the most relevant fragments were selected and probable chemical structures of these fragments were suggested employing CFM-ID code (Allen et al.,

2016). Proposed structures of some interesting *m/z* ions (n = 20) are shown in Table R3.4, while the other structures of *m/z* ions (n = 30) are shown in Supplementary material Table S3.4.

First, complete feasible fragments for candidate molecules (i.e. monosaccharides, POS-DP2 and DP3) were calculated by systematically breaking bonds within the molecule (Allen et al., 2016; Allen et al., 2015) in order to obtain a POS *in silico* fragmentation library. For monosaccharides, pectic neutral sugars (arabinose, rhamnose, xylose and galactose) and GalA (methyl- or acetyl-esterified or not) were considered. The POS-DP2 included, taking into account more possible structures present in HG and RG-I, were Di-GalA (methyl-esterified or not), xylose-α(1,3)-GalA (methyl- or acetyl-esterified or not), GalA-α(1,2)-rhamnose (acetylated or not), rhamnose-α(1,4)-GalA (acetylated or not), galactose-α(1,4)-rhamnose and arabinose-α(1,4)-rhamnose (Atmodjo et al., 2013; Mohnen, 2008).

Tri-POS library was generated considering combinations of all POS-DP2 structures and pectic monomers, following the main criteria reported by Atmodjo et al. (2013) and Mohnen (2008). Then, we proposed a chemical structure for each *m/z* ion by looking for these specific ruptures in our library for studying structural characteristics of known pectic sugars and unknown POS released during enzymatic hydrolysis. The abundances of several of the most relevant ions in machine learning classification were significantly higher ($p < 0.05$) in known pectic sugars released by the four enzymes, so fragments *m/z* 207 and 309 could be originated from all compounds studied. Ions *m/z* 311, 382, 414, 486, 498, 500, 501, 502, 513, 528, 532, 544, 545 and 546 were also higher in known pectic sugars and could also be derived from GalA-containing POS or GalA oligomers bonded by α(1,4) glycosidic linkages. Then, statistically significant structural differences among unknown POS-DP2 and POS-DP3 released with the four preparations were studied:

Pentopan. Some *m/z* ions were significantly more abundant in unknown POS present in these hydrolysates: ion *m/z* 237 which could be derived from all pectic sugars and POS and *m/z* 193, 425 and 530 originated from xylose and POS ruptures. Other more specific ions higher in these POS were *m/z* 488 originated from POS containing rhamnose-α(1,4)-GalA (acetylated); ion *m/z* 399 is formed from GalA-α(1,2)-rhamnose,

while *m/z* 529 is formed from POS containing rhamnose-α(1,4)-GalA (acetylated), GalA-α(1,2)-rhamnose (acetylated) or xylose-α(1,3)-GalA (methyl- or acetyl-esterified or not); and ion *m/z* 454 which could also be formed from POS containing arabinose-α(1,4)-rhamnose.

Glucanex. Some fragments with higher abundance in unknown POS from Glucanex, *m/z* 280 derived from rhamnose-α(1,4)-GalA (acetylated), *m/z* 370 derived from GalA-α(1,2)-rhamnose (acetylated), and *m/z* 369 which could be derived from acetylated dimers of rhamnose-α(1,4)-GalA or GalA-α(1,2)-rhamnose as well as GalA-containing POS bonded by α(1,4) glycosidic linkages. Other important ions in these POS were *m/z* 178 and 192 which could also be originated from the rupture of xylose-α(1,3)-GalA (acetylated). Interestingly, *m/z* 123 derived from POS containing GalA (methyl-esterified) was significantly higher in Glucanex POS and lower in those obtained with cellulase from *A. niger* perhaps this enzyme produced slight demethylation.

Pectinex Olio. The MS study of unknown POS obtained with Pectinex Olio revealed that the abundances of *m/z* 337, which could be originated from all pectic sugars and all POS, and *m/z* 340, 424 and 443, specifically originated from rhamnose-α(1,4)-GalA, GalA-α(1,2)-rhamnose or xylose-α(1,3)-GalA (methyl- or acetyl-esterified or not), were significantly higher in these hydrolysates.

Cellulase from A. niger. MS spectra of unknown POS formed with Cellulase from *A. niger* showed higher abundances of ions *m/z* 441, originated from galactose and all POS, and *m/z* 482, specifically derived from GalA-α(1,2)-rhamnose (acetylated) and xylose-α(1,3)-GalA (acetylated).

Table R3.4. Possible chemical structures of the most relevant *m/z* ions in POS classification given by machine learning algorithms, determined by CFM-ID. These ions correspond to TMSO fragments from pectic di- or trisaccharides.

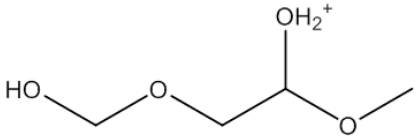
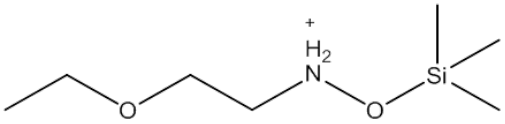
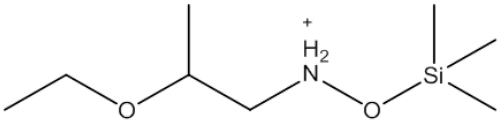
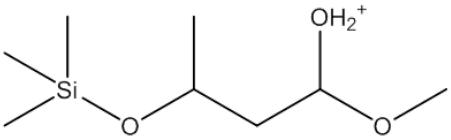
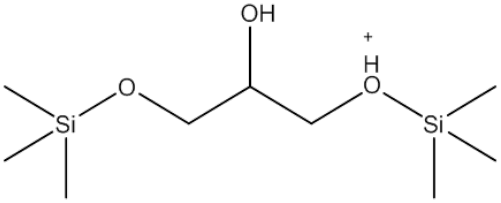
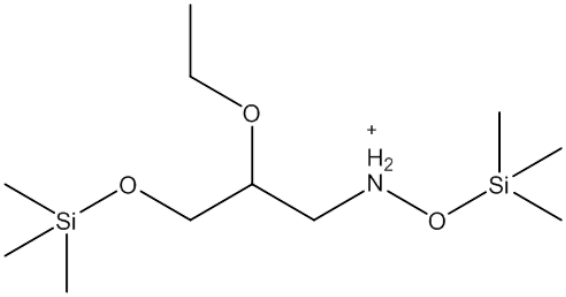
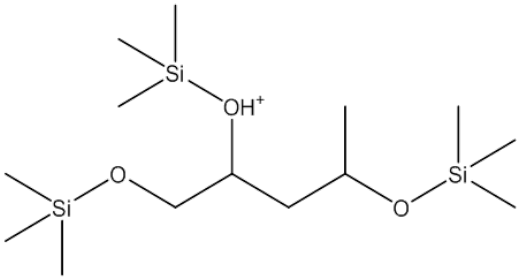
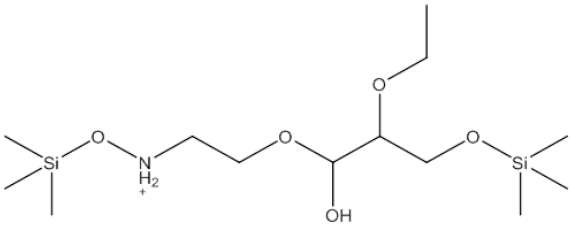
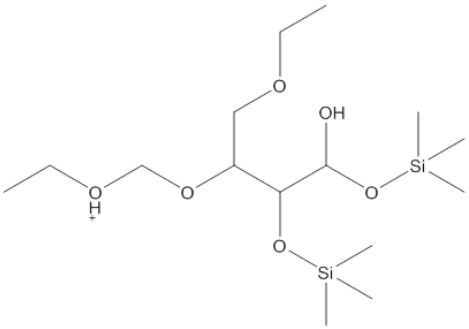
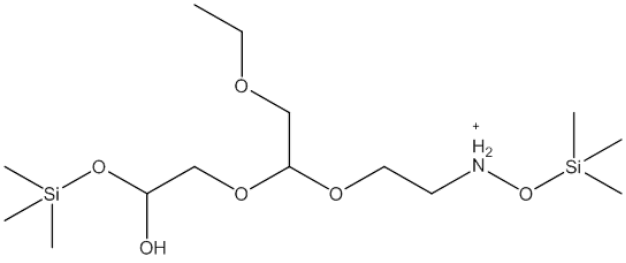
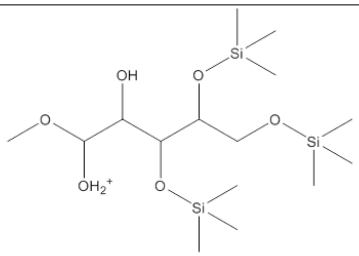
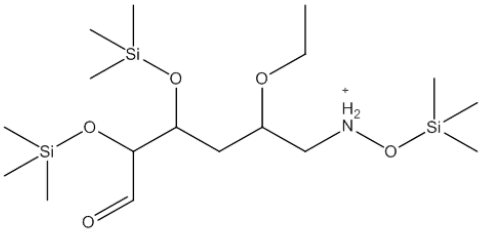
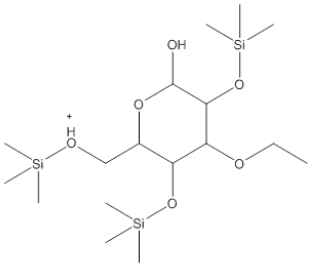
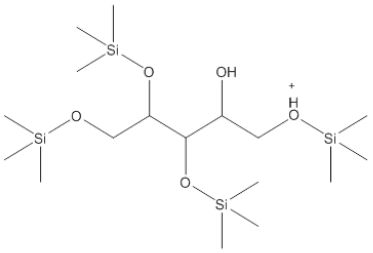
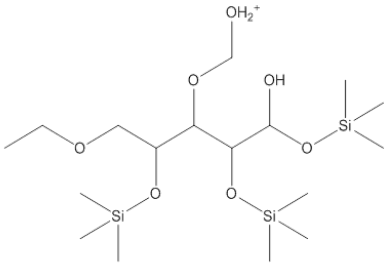
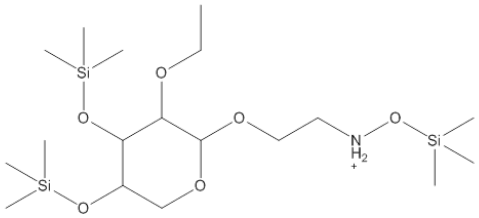
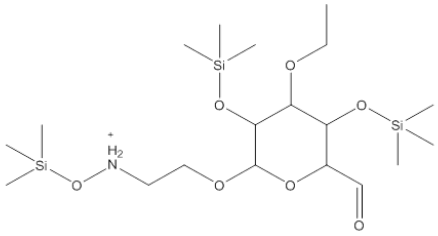
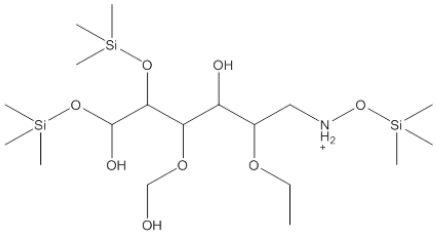
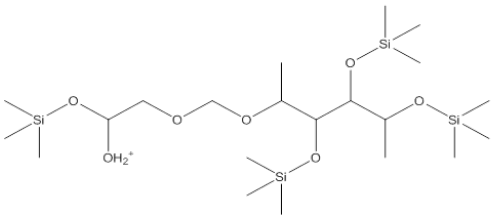
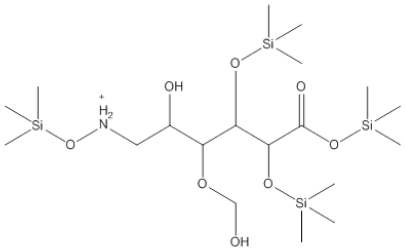
m/z	Structure	*m/z*	Structure
123		**178**	
192		**193**	
237		**280**	
337		**340**	
369		**370**	

Table R3.4. Cont.

m/z	Structure	*m/z*	Structure
399		**424**	
425		**441**	
443		**454**	
482		**488**	
529		**530**	

Other *m/z* ions which were relevant in more than one group were *m/z* 531 present in all POS containing GalA units and *m/z* 483 present in galactose-α(1,4)-rhamnose or GalA-α(1,2)-rhamnose. Similarly, abundances of *m/z* 353, 354, 457, 458, 471 and 472 were higher in unknown POS and could be derived from the rupture of most POS, with special importance for those containing rhamnose-α(1,4)-GalA, GalA-α(1,2)-rhamnose or xylose-α(1,3)-GalA. More specifically, *m/z* 225, 444, 469, 484 and 542, formed from the rupture of GalA and molecules containing GalA-α(1,4)-GalA dimers, were relevant in known pectic sugars and also in unknown POS from Pentopan and Glucanex preparations. Finally, *m/z* 439 formed from rhamnose-α(1,4)-GalA, GalA-α(1,2)-rhamnose and galactose-α(1,4)-rhamnose, was relevant in both hydrolysates with Pentopan or with Cellulase from *A. niger* preparations. At last, the dimer structures above referred may be also present in trisaccharides, but no trisaccharides specific *m/z* ions were relevant for POS classification. The assignment of these glycosidic linkages was possible due to the most frequent structures found in pectins (Atmodjo et al., 2013; Mohnen, 2008).

As expected, *m/z* fragments possibly formed from the rupture of dimers and trimers containing neutral sugars and GalA (methyl- or acetyl-esterified or not) units were relevant in unknown POS. In general terms, these results suggest that Glucanex and Pectinex Olio preparations produce POS that may contain rhamnose, xylose and GalA (methyl- or acetyl-esterified or not). In addition, arabinose may also be present in POS structures produced by Pentopan. Moreover, Cellulase from *A. niger* may produce POS that contain neutral sugars such as galactose, rhamnose and xylose, and also acetylated GalA units.

A GC–MS spectral study might be considered as a first approach to determine structural characteristics of POS enzymatically obtained using commercial preparations with different activities. MS data modelling may lead to a better understanding of differences observed in the chromatographic profiles of these samples. It has been demonstrated that is possible to classify complex oligosaccharides according to the enzyme used for their obtainment, by extracting chemically relevant information from their full spectra. These models could be applied on new reaction mixtures containing novel oligosaccharides to evaluate the activity of different enzyme preparations

considering their high prediction rates and allowing consistent structural information to be obtained. The ability to determine structural similarities and differences among novel POS can be of great importance to establish structure-function relationships which could be very useful to extrapolate results from biological assays.

CONCLUSIONS

Novel artichoke POS mixtures have been enzymatically obtained and characterized. Differences in the chromatographic profiles of POS were observed according to the enzyme used suggesting different structural properties, confirmed by several chromatographic techniques. Cellulase from *A. niger* and Pectinex® Ultra-Olio formed large amounts of POS, with M_w around 14 and 6.0 kDa, respectively while Glucanex produced high M_w POS and modified pectins, and Pentopan showed chromatographic profiles similar to those of original artichoke pectin. POS were analysed by HPAEC-PAD, showing that Cellulase from

A. niger preparation produced the highest amount (310.6 mg g^{-1} pectin). Mass spectra of unknown POS-DP2 and -DP3 have been studied and classified using machine learning algorithms (multilayer perceptron, random forest and boosted logistic regression) obtaining high prediction rates on the test set. These models confirm structural differences observed in the hydrolysis profiles of commercial preparations with various enzymatic activities and could be used to establish structure-function relationships.

Article IV

6.2.2. Article IV: Enzymatic Production and Characterization of Pectic Oligosaccharides Derived from Citrus and Apple Pectins: A GC-MS Study Using Random Forests and Association Rule Learning

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

Cite This: J. Agric. Food Chem. 2019, 67, 7435−7447

pubs.acs.org/JAFC

Enzymatic Production and Characterization of Pectic Oligosaccharides Derived from Citrus and Apple Pectins: A GC-MS Study Using Random Forests and Association Rule Learning

Carlos Sabater, Alvaro Ferreira-Lazarte, Antonia Montilla, and Nieves Corzo

Institute of Food Science Research, CIAL (CSIC-UAM), CEI (UAM+CSIC), C/NicolaśCabrera 9, Madrid 28049, Spain

Link: https://pubs.acs.org/doi/abs/10.1021/acs.jafc.9b00930

A B S T R A C T

Pectic oligosaccharides (POS) from citrus and apple pectin hydrolysis using ViscozymeL and Glucanex200G have been obtained. According to the results, maximum POS formation was achieved from citrus pectin after 30 min of hydrolysis with ViscozymeL, with a yield of 652 mg g^{-1} and average molecular weight (M_w) of 0.8−2.5 kDa, while with Glucanex200G, the yield was 518 mg g^{-1} and M_w was 0.8−7.1 kDa. Digalacturonic and trigalacturonic acids were identified among other low M_w compounds as di- and tri-POS. In addition, differences in GC-MS spectra of all oligosaccharides found in the hydrolysates were studied by employing random forests

and other algorithms to identify structural differences between the obtained POS, and high prediction rates were shown for new samples. Chemical structures were proposed for some influential *m*/*z* ions, and 12 association rules that explain differences according to pectin and enzyme origin were built. This information could be used to establish structure−function relationships of POS.

Keywords: Pectic oligosaccharides, ViscozymeL, Glucanex200G, Machine learning, *in silico* fragmentation

INTRODUCTION

Currently, there is an overall interest in obtaining novel oligosaccharides with improved functional properties using plant cell wall polysaccharides as raw material (de Moura et al., 2015). Pectic oligosaccharides (POS) have been proposed as a new class of prebiotics due to their health-promoting activities including prevention of the adhesion of uropathogenic microorganisms, stimulation of apoptosis of human colonic adenocarcinoma cells, or cardiovascular protection (Holck et al., 2014).

POS can be obtained from depolymerization of raw materials or purified pectins mainly extracted from apple, citrus peel, and sugar beet byproducts, using enzymes, chemical reagents (acids), or combined methods (Gómez et al., 2013; Martínez et al., 2009a,b). Also, other substrates, such as onion skins (Baldassarre et al., 2018), rapeseed cake (Cobs-Rosas et al., 2015), and artichoke pectin (Sabater et al., 2019a), have been used to obtain POS. On the other hand, pectins are one of the most structurally complex families of polysaccharides in nature containing three major pectic polysaccharides: homogalacturonan (HG) and rhamnogalacturonan I and II (RG-I and RG-II). HG, the most abundant domain, is a homopolymer of $_{D}$-galacturonic acid (GalA) with α-1,4 linkages (GalA ∼ 65% of pectin). This polymer is partially methyl-esterified at the *O*-6 position and may be *O*-acetylated at *O*-2 or *O*-3; also, some GalA units may be substituted at *O*-3 with β-$_{D}$-xylose, forming xylogalacturonan (XGA). The second most important domain is RG-I, with alternate sequences of GalA and $_{L}$-rhamnosyl (Rha) residues in its backbone [-4)-α-$_{D}$-Gal*p*A-(1,2)-α-$_{L}$-Rha*p*-(1-]$_{n}$. In this domain, rhamnose may be substituted at *O*-4 with linear or branched oligosaccharides of arabinose

(arabinans), galactose (galactans), or both (arabinogalactans). The extent of branching depends on neutral sugar content, so it may vary according to pectin source. RG-II, consisting of up to 12 different types of sugars in over 20 different linkages, constitutes another minor domain of pectin (Caffall & Mohnen, 2009). Due to complexity of pectin structure, the resulting POS can have varied chemical structure, which will influence their bioactive properties.

To date, most studied food-grade POS have been produced by commercial enzymes, mostly pectinases, usually produced by *Aspergillus* spp (Ma et al., 2018; Delattre et al., 2008). These POS can present very different chemical structures depending on the type of enzyme, reaction conditions, and substrate used (de Moura et al., 2015).

Different analytical techniques are employed to identify and characterize POS: high performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) (Sabater et al., 2019a; Gómez et al., 2013), high performance size-exclusion liquid chromatography with refractive index detector (HPSEC-RID) (Muñoz-Almagro et al., 2017), high performance size-exclusion liquid chromatography with evaporative light scattering detector (HPSEC-ELSD) (Sabater et al., 2019a), and hydrophilic interaction liquid chromatography-mass spectrometry (HILIC-MS) (Leijdekkers et al., 2015), the latter being one of the most used. However, GC-MS can be of great interest because it generates massive amounts of high-dimensional data (Sabater et al., 2019a). To interpret GC-EI-MS results, *in silico* fragmentation methods have been developed (Allen et al., 2016; Allen et al., 2015).

Another important step in MS data analysis involves data modeling to find valuable information about chemical structures. With this aim, several machine learning algorithms have been applied (Yi et al., 2016). Some common algorithms for classifying metabolomics and food science data include random forests (RF) and artificial neural networks (ANN) (Sabater et al., 2018b; Lim et al., 2017; Moncayo et al., 2017). These models could be of great importance for sample characterization. ANN have been applied on GC-MS data to classify small molecular structures (Gosav et al., 2011) and compared to RF and boosted logistic regression (BLR) to characterize POS substructures obtained from low methyl-esterified artichoke pectin (Sabater et al., 2019a). However, more studies are needed for determining GC-MS structural differences of POS, for example, POS obtained from an HG type structure and highly

methyl-esterified pectin. In addition, in a few studies, association rules between *m*/*z* ions for establishing fragmentation patterns that could explain these differences have been calculated.

Therefore, the aim of this work was to characterize novel POS obtained from the hydrolysis of polygalacturonic acid (PGA) and highly methyl-esterified pectin (from citrus and apple) using two commercial enzymes with different activities, endopolygalacturonase (endo-PG) and others, including arabinosidase and galactosidase, through their GC-MS spectra. For this purpose, spectral data modeling has been improved to gain a better understanding of oligosaccharide fragment relationships. Machine learning algorithms (RF, BLR, and ANN) were computed, and most relevant fragments were selected to build association rules that explain some fragmentation patterns and structural differences of POS depending on the enzyme and sample types.

MATERIALS AND METHODS

Standards and Reagents

Analytical reference substances such as sucrose, $_{L}$-arabinose, $_{D}$-xylose, $_{L}$-rhamnose, $_{D}$-galactose, $_{D}$-mannose, D-glucose, $_{D}$-fructose, sucrose, kestose, nystose, galacturonic acid (GalA), digalacturonic acid (Di-GalA), PGA (purity >85%, GalA > 96%, degree of methyl-esterification (DM) of 82.0 ± 0.0%, determined by FT-IR), and phenyl-β-glucoside were purchased from Sigma-Aldrich (Steinheim, Germany). Trigalacturonic acid (Tri-GalA) was from Carbosynth (Compton, UK). Commercial apple pectin with a GalA > 74%, according to the label, and DM of 83.0 ± 0.5% was acquired from Sigma-Aldrich. Commercial citrus pectin with a GalA > 74% and a molecular weight (M_w) ranging from 22 to 400 kDa, according to the label, and DM of 71.0 ± 1.4% was acquired from Acofarma (Terrassa, Barcelona). Commercial enzyme preparations Glucanex200G (Glucanex, β-glucanase from *Trichoderma harzianum*) and ViscozymeL (Viscozyme, endo-1,3(4)-β-glucanase from *A. aculeatus*) were a generous gift from Novozymes.

Enzyme Characterization

Enzyme characterization assays were carried out using the method of DNS with a 1% (w/v) PGA solution dissolved in 0.05 M acetate buffer (pH 4.5) as substrate and 1 mg mL^{-1} of Glucanex or 1 μL mL^{-1} of Viscozyme. Samples were incubated at 50 °C for 2 min. One unit of pectinase activity was defined as the amount of enzyme required to release 1 μmol of GalA per min at 50 °C. Pectinase activity was 3778 U mL^{-1} and 739 U g^{-1} for Viscozyme and Glucanex preparations, respectively.

Total protein was determined by the Bradford method as 203 mg mL^{-1} and 172 mg g^{-1} for Viscozyme and Glucanex, respectively. Specific pectinase activity was 18.6 and 4.3 U mg^{-1} protein for Viscozyme and Glucanex preparations, respectively.

Pectin Hydrolysis

Two pectin depolymerization experiments were carried out (Figure S4.1).

Hydrolysis of Pectins To Know Monosaccharide Composition. Citrus and apple pectins were submitted to a chemical and enzymatic hydrolysis.

For the chemical hydrolysis, 10 mg of pectin was hydrolyzed with 500 μL of 2 M trifluoroacetic acid (TFA) at 110 °C under inert conditions for 4 h (Sabater et al., 2018a).

For the enzymatic hydrolysis, 1% (w/v) commercial pectin solutions in 0.05 M acetate buffer (pH 4.5) were hydrolyzed with 90 U mL^{-1} of Viscozyme. Samples were incubated in a thermomixer at 50 °C and 750 rpm for 24 h (Sabater et al., 2018a).

Enzymatic Obtainment of Pectic Oligosaccharides (POS)

POS were obtained by enzymatic hydrolysis of 1% (w/v) commercial pectin solutions and PGA, prepared in 0.05 M acetate buffer (pH 4.5), using 16 and 4 U mL^{-1} of Viscozyme and 16 U mL^{-1} of Glucanex incubated in an orbital shaker at 50 °C and 750 rpm.

Aliquots (1 mL) were withdrawn from the reaction mixture at short (30 min) and long times (24 h) and immediately immersed in boiling water for 5 min to inactivate the enzyme. Samples were stored at −18 °C for subsequent analysis. Enzymatic reactions were carried out in duplicate, and analyses were repeated at least twice for each enzymatic treatment.

Chromatographic Determination of Carbohydrates

Gas Chromatography with Flame Ionization Detector (GC-FID) and Mass Spectrometry (GC-MS)

Carbohydrates were analyzed as trimethylsilylated oximes (TMSO). GC-FID and GC-MS analyses were performed according to the method described by Sabater et al. (2019a).

High Performance Size-Exclusion Liquid Chromatography with Refractive Index Detector (HPSEC-RID).

The M_w distribution of carbohydrates in the reaction mixtures was determined by HPSEC-RID according to Muñoz-Almagro et al. (2017).

High Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD)

POS formed after 30 min of enzymatic hydrolysis were analyzed by HPAEC-PAD following the method of Sabater et al. (2019a).

Statistical Analyses

ANOVA tests and Tukey's test for $p < 0.05$ were applied to all data generated.

Study of GC-MS Spectra

In order to study the structural differences between POS (di- and trisaccharides) obtained with enzymatic preparations, a comprehensive GC-MS study was carried out (Figure R4.1). Full spectra (n = 212) of unknown POS and known pectic sugars (arabinose, rhamnose, xylose, galactose, GalA, Di-GalA, and Tri-GalA) present in reaction mixtures (Table S4.1) were analyzed and classified according to three different criteria: (i) pectin type, (ii) enzyme type, and (iii) pectin and enzyme types. For this purpose, three machine learning models were used and compared to reinforce the results from each individual classification: random forests (RF), boosted logistic regression (BLR), and artificial neural networks (multilayer perceptron, MLP).

Signal Processing. Discrete wavelet transform (DWT) was applied to reduce spectral noise and, subsequently, increase model performance according to previous studies (Sabater et al. 2019a). A 14 level DWT was computed. In the threshold step, DWT coefficients under the 15% percentile were removed to denoise the signal.

Classification Studies. Known pectic sugars and unknown POS preprocessed GC-MS spectra were introduced as inputs in three classification studies. RF and BLR were selected for the first two (i, ii) while MLP was also included in the third one (iii) to compare the predictive power. The following hyperparameters were tuned to compute the models:

RF: A RF model was built with 500 trees and 2 variables tried at each split for the first study while two RF with 1000 trees and 2 variables tried at each split were computed for the rest.

BLR: A BLR consisting of 21 iterations was computed for all studies.

MLP: A MLP was built with 1 hidden layer consisting of 41 hidden neurons. The activation function was logistic.

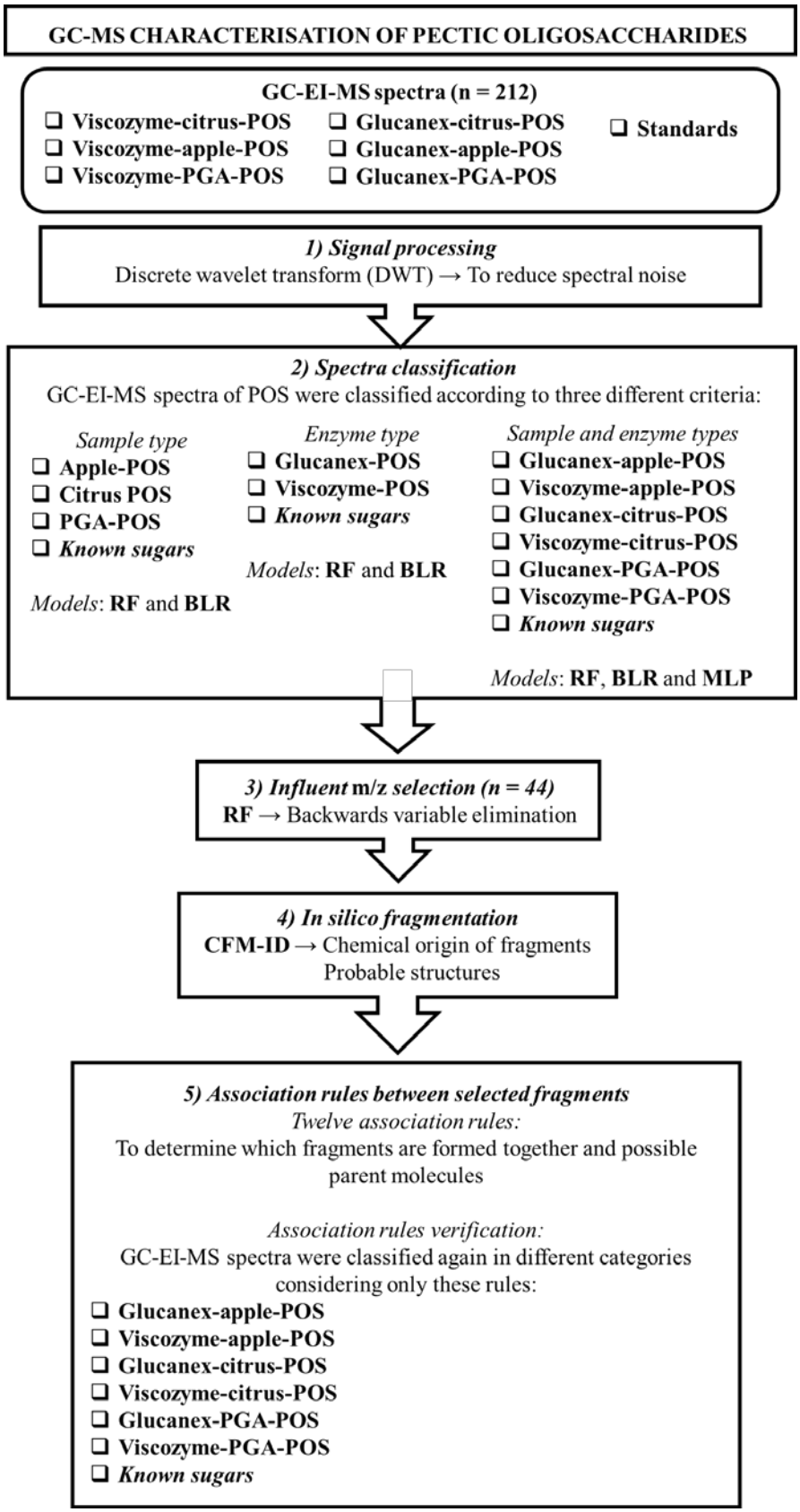


Figure R4.1. Schematic representation of the GC-EI-MS computational study.

All the models were trained with 70% of the data, 10-fold cross-validated, and then tested with 30% of the data from each class (corresponding to new samples). Predictive power of machine learning algorithms was also compared to principal components analysis (PCA) and partial least-squares-discriminant analysis (PLS-DA). Variable importance analysis was carried out to determine the most influential *m*/*z* ions in each model according to Sabater et al. (2019a).

Influential m/z Selection. Most relevant and non-redundant *m*/*z* ions in each classification were selected by employing RF models for backward variable elimination. The least important variables were successively eliminated using the out of bag (OOB) error as the minimization criterion.

In Silico Fragmentation. After determining the most important *m*/*z* ions in POS classification, chemical structures of some fragments have been proposed by employing the competitive fragmentation modeling source code (CFM-ID) developed by Allen et al. (2016). GC-EI-MS *in silico* fragmentation patterns of POS structures, described in the bibliography (Atmodjo et al., 2013; Caffall & Mohnen, 2009), were determined and compared to those of enzymatically obtained POS (Sabater et al., 2019a).

Association Rules between Fragments. Association rule-based classification of GC-MS spectra considering relevant fragments selected by RF was carried out. Twelve association rules that could explain differences in the fragmentation pattern of POS from different enzymatic and pectin origin were established. This model was trained with 70% of the data and tested with 30% of new samples.

All statistical analyses were computed on R v3.5.0. DWT was performed using the wavelets package (Aldrich, 2013). MLP was computed using the RSNNS package (Bergmeir & Benitez, 2012), and RF classification was performed with the Random Forest package (Liaw & Wiener, 2002). For BLR, the caTools package was employed (Tuszynski, 2014). Association rules were computed using the arulesCBA package (Johnson & Hahsler, 2018).

RESULTS AND DISCUSSION

Monomeric Characterization of Pectin

The monosaccharide composition of commercial citrus and apple pectins was determined by GC-FID in acid and enzymatic hydrolysates; Figure S4.2 shows a chromatogram of the carbohydrates found in xylose, arabinose, rhamnose, galactose, mannose, glucose, and GalA. The quantification of pectic monosaccharides is shown in Table R4.1. The results showed that chemical hydrolysis of pectin at high temperatures may be chosen in order to quantify neutral sugars, as it has been suggested by Soria et al. (2014) because the enzymatic hydrolysis did not lead to a complete release of arabinose and rhamnose. Gama et al. (2015) determined a low α-$_L$-arabinofuranosidase activity in Viscozyme during enzymatic hydrolysis of apple pomace. However, the enzymatic hydrolysis was most appropriate for quantifying GalA due to its thermolabile nature. Thus, enzymatic hydrolysis of pectin using the Viscozyme preparation at long times (24 h) and acid hydrolysis (4 h) allowed one to determine the GalA and neutral sugars, respectively, providing a better characterization of pectin (Sabater et al., 2018a). Other authors also have determined the GalA and neutral sugars in enzymatic and acid hydrolysates, respectively (Robert et al., 2008).

Taking into account the data shown in Table R4.1, the GalA content was the most abundant in both pectins, corresponding to 89.6% and 89.3% for apple and citrus pectin, respectively. Considering acid hydrolysis data, the former contained more neutral monosaccharides in its structure, i.e., arabinose and rhamnose, depicting the relevance of rhamnogalacturonan (RG) structures. However, xylose contents obtained with these two treatments were similar. The presence of high amounts of arabinose and xylose was also reported in apple pectin by the larger presence of xylo-galacturonan and arabino-galactans (Caffall & Mohnen, 2009). Glucose and mannose are not pectic carbohydrates, and they were present due to a possible extraction from cellulose and hemicelluloses including mannans during pectin obtainment.

Production of Pectic Oligosaccharides (POS) derived from Citrus and Apple Pectins

Considering that enzymatic hydrolysis is one of the best methodologies for obtaining POS, in this study, two commercial enzyme preparations were compared, Viscozyme and Glucanex with high and low polygalacturonase activity, respectively. Depolymerization of citrus and apple pectins and PGA (used as reference substrate) was carried out using Viscozyme (4 and 16 U mL^{-1}) and Glucanex (16 U mL^{-1}). Production of POS was followed by HPSEC-RID. When enzymatic hydrolysis of citrus and apple pectins were carried out using 16 U mL^{-1} of Viscozyme even after 30 min of reaction, a high hydrolysis rate was found and most oligosaccharides (>80%) had an M_w below 1 kDa.

Table R4.1. Monosaccharide content (mg g^{-1} of pectin) after acid (TFA 2 M, 110 °C, 4 h) and enzymatic (Viscozyme®L 90 U mL^{-1}, pH 4.5, 50 ºC, 24 h) hydrolysis of citrus and apple pectins.

Carbohydrates (mg g^{-1} pectin)	**Citrus pectin**		**Apple pectin**	
	Acid Hydrolysis	***Enzymatic Hydrolysis***	***Acid Hydrolysis***	***Enzymatic Hydrolysis***
Xylose	0.9 ± 0.1 [b]	1.0 ± 0.3 [b]	5.6 ± 2.1 [a]	3.4 ± 2.1 [a, b]
Arabinose	27.1 ± 0.9 [b]	7.3 ± 0.6 [c]	65.7 ± 6.8 [a]	21.9 ± 0.6 [b]
Rhamnose (Rh)	21.1 ± 0.1 [a]	14.8 ± 0.4 [b]	23.9 ± 2.4 [a]	13.5 ± 0.4 [b]
Galactose	69.5 ± 5.5 [a]	82.0 ± 3.5 [a]	53.1 ± 2.7 [b]	43.4 ± 1.7 [b]
Galacturonic acid (GalA)	656.1 ± 35.6 [a, b]	875.0 ± 24.2 [a]	631.3 ± 97.5 [b]	704.5 ± 31.6 [a, b]
GalA/Rh	31 [c]	59 [a]	26 [d]	52 [b]

[a, b, c, d] Statistically significant differences between groups.

The chromatographic profiles of hydrolyzed citrus and apple pectin using Viscozyme (4 U mL^{-1}) and Glucanex (16 U mL^{-1}) are shown in Figure R4.2. As it can be seen, the molecular weight average (M_w) distributions of hydrolysates were very different. Samples hydrolyzed with Glucanex (red line) showed a wider variety of fragments and

higher M_w than those treated with Viscozyme (blue line). In contrast, chromatographic profiles of hydrolyzed citrus and apple pectin with Viscozyme were similar, despite differences in the monomer composition and the initial M_w of both pectins, 547 and 427 kDa for citrus and apple pectin, respectively, and 169 kDa for PGA. Similar profiles were also obtained when citrus and apple pectin were hydrolyzed with Glucanex.

Figure R4.3 shows the distribution of M_w of the different fragments (POS) produced and their concentrations (mg g^{-1} pectin) in hydrolysates at 30 min and 24 h using Viscozyme (4 U mL^{-1}) and Glucanex (16 U mL^{-1}). As already stated, similar hydrolysis patterns were obtained for the two pectins using Viscozyme, and the same behavior was observed when pectins were hydrolyzed with Glucanex. With respect to PGA, the behavior was similar when hydrolysis was performed using Glucanex, while PGA was more quickly hydrolyzed with Viscozyme.

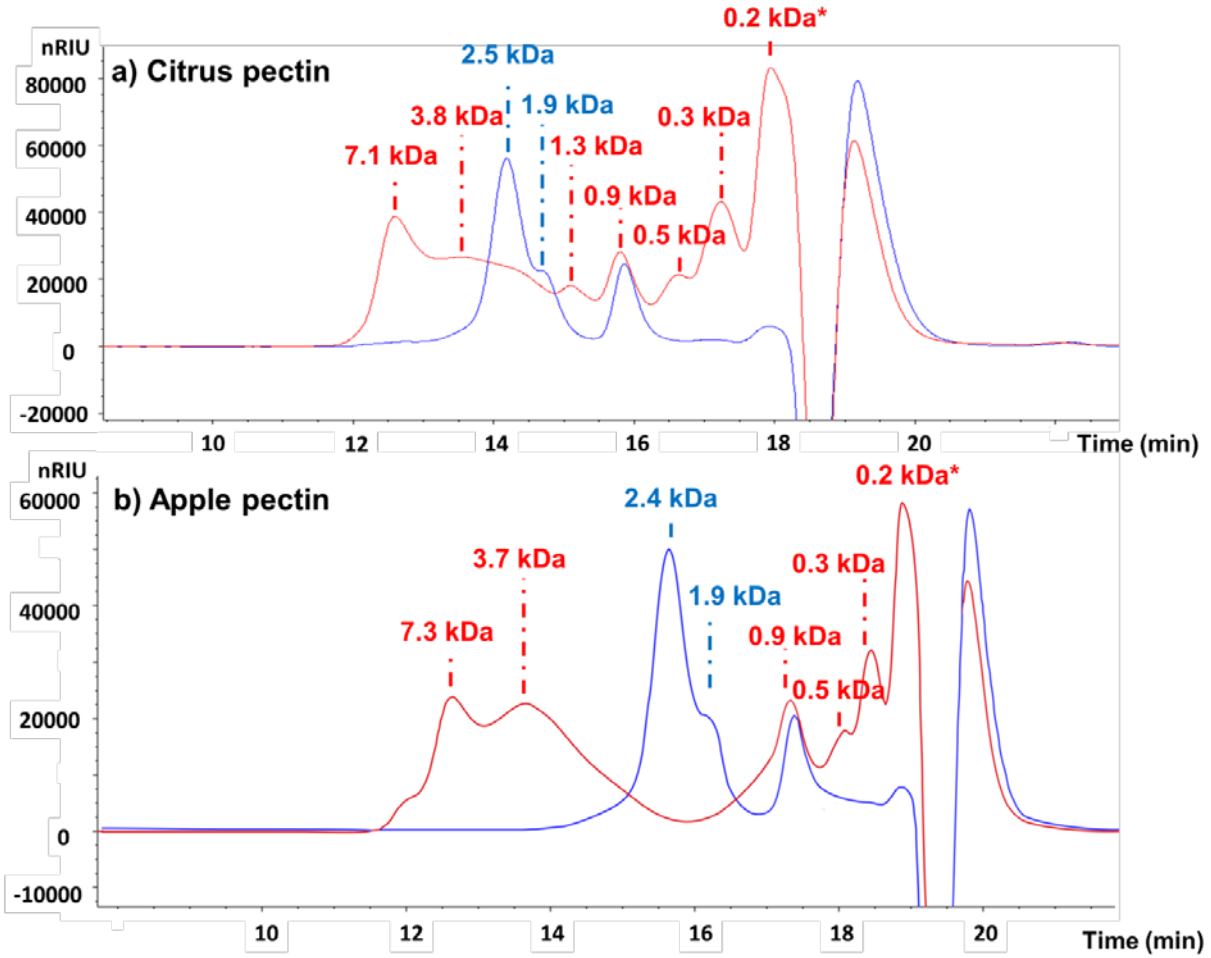


Figure R4.2. HPSEC-RID profiles of hydrolyzed citrus (a) and apple (b) pectins (30 min at 50 °C) using 16 U mL^{-1} of Glucanex (red) and 4 U mL^{-1} of Viscozyme (blue) preparations. *Compound (0.2 kDa) present in the Glucanex preparation.

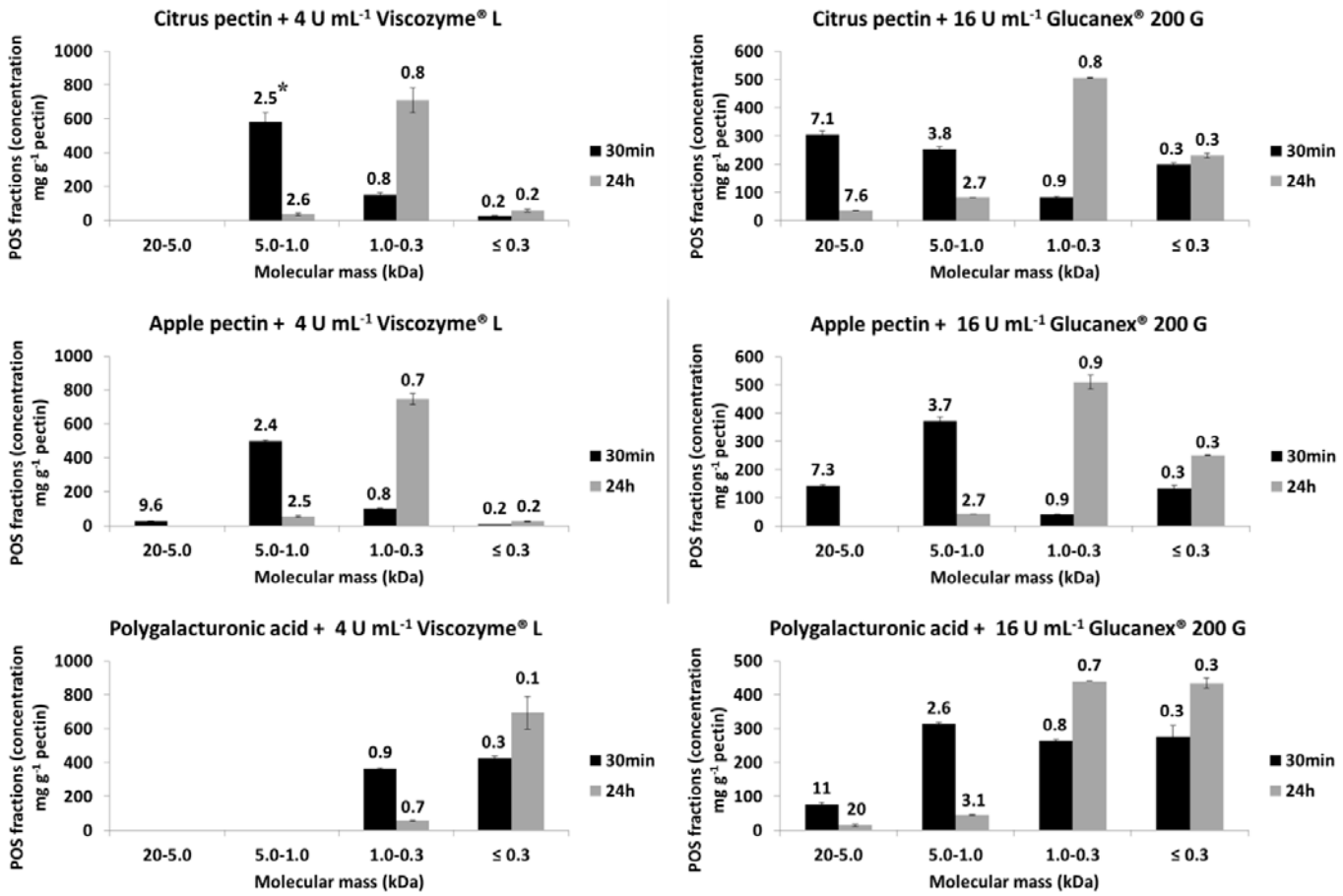


Figure R4.3. Molecular weight distribution of POS (HPSEC-RID) obtained from hydrolysates of commercial citrus and apple pectins and polygalacturonic acid (as pectin standard) using Glucanex (16 U mL^{-1}) and Viscozyme (4 U mL^{-1}) for 30 min and 24 h. *Average molecular weight of each region.

The quantitative data of low M_w compounds released during pectin hydrolysis showed those differences between enzymes. After 30 min of incubation with Glucanex, POS with a M_w of around of 7−11 kDa comprized 36.3%, 20.7%, and 8.3% of total POS obtained from citrus and apple pectin and PGA, respectively, and fragments of around 2.6−3.8 kDa were approximately 30.1%, 53.9%, and 33.7% for the same samples. Also, POS with M_w from 0.8 to 0.9 and di- and monosaccharides (0.2−0.3 kDa) were determined in assayed pectins and PGA. After 24 h of hydrolysis, the release of low M_w (<1 kDa) compounds was very high.

In contrast, after 30 min of enzymatic hydrolysis with Viscozyme, most obtained POS had a M_w of 2.4−2.5 kDa, comprized of around 76.4% and 78.7% for citrus and apple pectin. After 24 h of reaction, >80% of the fragments were below 1 kDa in all hydrolysates.

These results corroborated the importance of enzyme type in the formation of POS. Several studies reported the influence of enzyme and reaction conditions over POS formation (Babbar et al., 2016a; Combo et al., 2012). Pectin source is another influential factor for the type and yield of POS formed; thus, enzymatic hydrolysis of highly methyl-esterified and low methyl-esterified pectin by endopolygalacturonase (endo-PG) that randomly hydrolyses the homogalacturonan backbone yielded 95% POS with an average M_w of 3.5 and 3.0 kDa, respectively (Combo et al., 2012). Thus, POS with a similar M_w may be obtained by using pectin from different sources (Olano-Martin et al., 2001). A mixture of four endoenzymes (pectin methyl esterase, endo-PG, endoarabinosidase, and endogalactosidase) has been used in order to obtain POS of around 3 kDa from sugar beet pectin, predominantly 0.9 kDa oligogalacturonides (Ralet et al., 2005).

In order to know possible structures of POS present in hydrolysates of citrus and apple pectins using Glucanex or Viscozyme, HPAEC-PAD analysis was carried out; Figure R4.4 shows the chromatographic profiles of carbohydrates found. Acidic sugars, GalA (peak 1), Di-GalA (peak 2), and Tri-GalA (peak 3) eluted after neutral sugars. As with HPSEC-RID profiles of pectins, differences in POS profiles depending on the enzyme preparation were observed. In Glucanex (blue line) hydrolysates, a wide variety of acidic peaks with a degree of polymerization (DP) of 2, 3, and >3 were detected. When the retention times were compared with standards of GalA and di- and tri-GalA,

these compounds could be considered oligosaccharides with one molecule of GalA and one or more molecules of neutral sugars. This different pattern could highlight a different behavior of these enzymes; thus, Viscozyme could have an important endo-PG activity while Glucanex could have other glycosidase activities. All POS formed were quantified (Table R4.2). It can be observed that Viscozyme releases higher amounts of Di-GalA and Tri-GalA than Glucanex, while this enzyme released a great amount of disaccharides with GalA. HPAEC-PAD profiles showed that high mono-, di-, and tri-GalA levels were detected. According to the results obtained by HPAEC-PAD, POS formation from apple and citrus pectin after 30 min of hydrolysis ranged from 335 to 652 mg g^{-1} pectin. The highest formation of POS (767 mg g^{-1}) was obtained with Viscozyme preparation using standard PGA as substrate.

Other authors reported the use of Viscozyme to obtain POS from different origins, either directly from pectin-rich by-products, such as lemon peel, orange peel, and sugar beet pulp (Gómez et al., 2013; Martinez Sabajanes et al., 2012; Martínez et al., 2009a,b), or pectin isolated from sugar beet pulp, onion skins, rapeseed cake, and PGA (Baldassarre et al., 2018; Elst et al., 2018; Cobs-Rosas et al., 2015; Combo et al., 2012). Our results agree with these studies where high amounts of low M_w POS were obtained. On the other hand, Glucanex has been employed in the digestion of protoplasts from fungal spores (Cheng et al., 2000) and to obtain POS and modified pectin from artichoke pectin (Sabater et al., 2019a). In this previous study, a low concentration of enzyme was used (0.63 U mL^{-1}); modified pectin with a M_w of 61 kDa was obtained, determined by HPSEC-ELSD, and a very low amount of POS, 63.7 mg g^{-1}, was quantified by HPAEC-PAD.

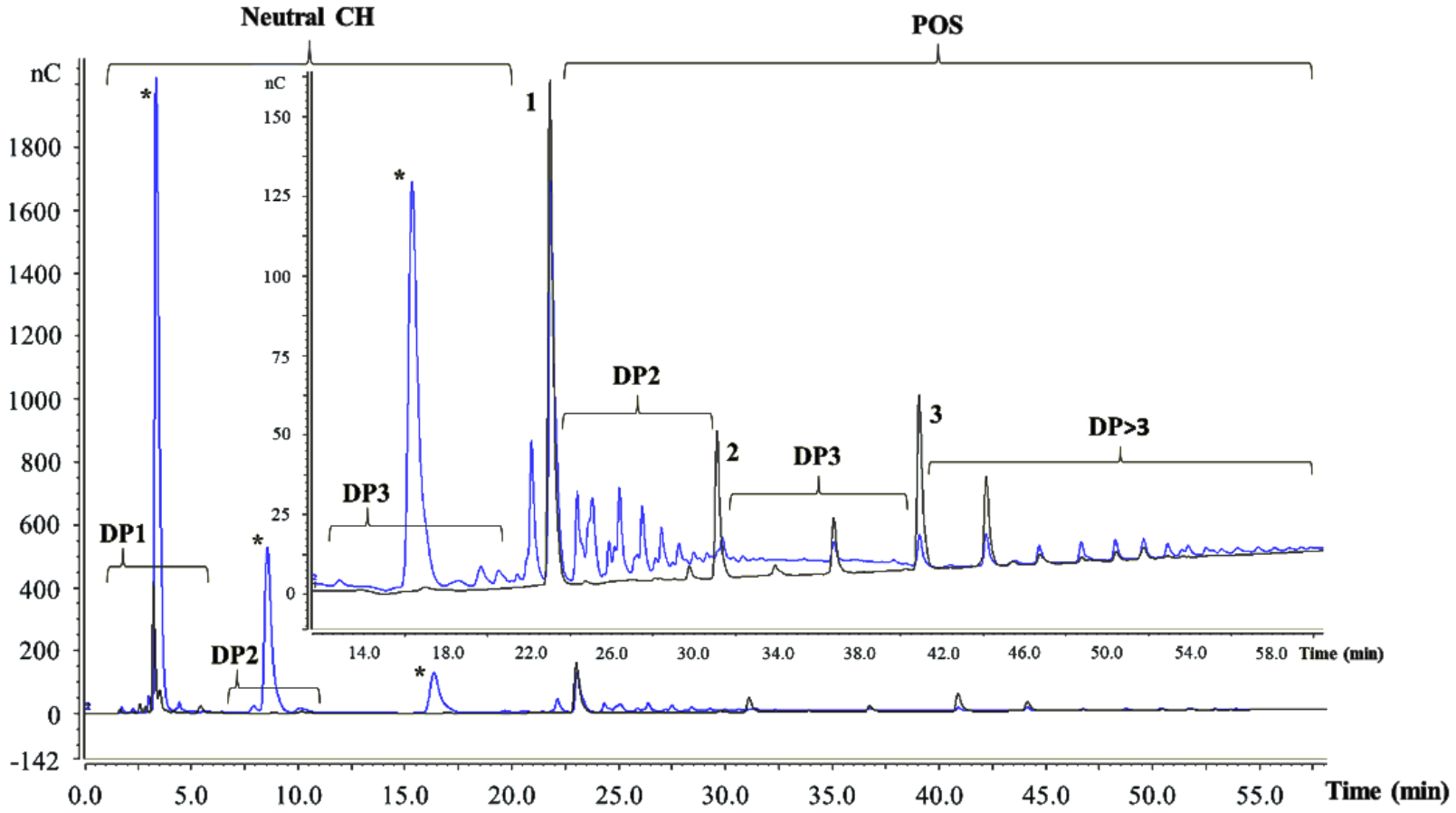


Figure R4.4. HPAEC-PAD profiles of carbohydrates found in hydrolyzed citrus pectin with 16 U mL^{-1} Glucanex200G (blue) and 4 U mL^{-1} Viscozyme (black) preparations for 30 min. DP: degree of polymerization; CH: carbohydrates; POS: pectic oligosaccharides. Peaks: (1) galacturonic acid, (2) digalacturonic acid, and (3) trigalacturonic acid. (*) Neutral carbohydrates present in the Glucanex200G preparation were not quantified.

Table R4.2. HPAEC-PAD quantification (mg g^{-1} pectin) of pectic oligosaccharides (POS) in hydrolysates (reaction time: 30 min) of studied pectins and PGA (polygalacturonic acid). GalA: galacturonic acid, Di-GalA: digalacturonic acid, Tri-GalA: trigalacturonic acid, POS: pectic oligosaccharides.

		POS (mg g^{-1} pectin)						
Enzymes	**Pectin**	**GalA (peak4)**	**Disaccharides with GalA**	**Di-GalA (peak5)**	**Trisaccharides with GalA**	**Tri-GalA (peak6)**	**Tetrasacchrides (or >3) wtih GalA**	**Total POS***
Viscozyme®L (4 U mL^{-})	**Citrus**	123 ± 5 ^{c}	-	166 ± 12 ^{c}	21 ± 1 ^{b}	67 ± 5 ^{c}	399 ± 39 ^{a}	652 ± 46 $^{a, b}$
	Apple	92 ± 7 $^{c, d}$	-	125 ± 4 ^{c}	19 ± 1 ^{b}	52 ± 4 $^{c, d}$	321 ± 23 ^{a}	517 ± 51 $^{b, c}$
	PGA	200 ± 8 ^{b}	-	352 ± 25 ^{a}	21 ± 1 ^{b}	338 ± 24 ^{a}	56 ± 6 ^{d}	767 ± 54 ^{a}
Glucanex®200G (16 U mL^{-1})	**Citrus**	93 ± 7 $^{c, d}$	238 ± 10 ^{a}	29 ± 1 ^{d}	16 ± 1 ^{b}	18 ± 1 ^{d}	217 ± 15 ^{b}	518 ± 51 $^{b, c}$
	Apple	60 ± 3 ^{d}	152 ± 11 ^{b}	19 ± 1 ^{d}	13 ± 0 ^{b}	11 ± 1 ^{d}	139 ± 14 $^{b, c}$	335 ± 24 ^{c}
	PGA	319 ± 23 ^{a}	18 ± 1 ^{c}	247 ± 7 ^{b}	75 ± 5 ^{a}	190 ± 13 ^{b}	111 ± 8 $^{c, d}$	641 ± 63 $^{a, b}$

* Total POS: Σ DP2, Di-GalA, DP3, Tri-GalA and DP>3.

$^{a, b, c, d}$ Statistically significant differences between groups.

Characterization of POS Present in Reaction Mixtures

To gain a deeper knowledge of the structure of POS released during enzymatic hydrolysis of apple and citrus pectins and PGA, GC-MS analyses of hydrolysates were carried out. Neutral monosaccharides, GalA, Di-GalA, and Tri-GalA as well as unknown di- and trisaccharides, perhaps corresponding to compounds eluted in HPAEC-PAD profiles between GalA and Di-GalA (Figure R4.4), were detected in the GC-MS profiles of citrus and apple pectin and PGA hydrolyzed with Glucanex (Figure S4.3a). Moreover, *m*/*z* ions 277, 321, 332, 333, 423, and 540 were present in the MS spectra of Di-GalA (Figure S4.3b). As previously indicated by Füzfai et al. (2004), *m*/*z* 332 and 540 were characteristic ions from GalA. In addition, *m*/*z* ions 277, 321, 333, and 423, derived from β-cleavage fragmentation of uronic acids, were reported in small abundance (Peterson, 1974).

GC-MS Spectra Classification Using Supervized Machine Learning Algorithms

Due to the scarce information found in the bibliography about the GC-MS spectra of these type of compounds, analysis of MS spectra employing machine learning algorithms (RF and BLR) was performed using full mass spectra of all disaccharides and trisaccharides found in the enzymatic hydrolysates, except Di- and Tri-GalA (identified with standards and their structure confirmed by GC-MS); Table S4.1 shows the number of spectra used to perform this study.

Before data analysis, each MS spectrum was decomposed and reconstructed using the DWT. Models were built using ions whose abundances were statistically different among groups ($p < 0.05$) and might correspond to known POS ruptures, assessed by competitive fragmentation modeling (CFM-ID) (Allen et al., 2016), as inputs.

Unknown POS were classified according to three different criteria: (i) sample type (apple pectin, $n = 40$; citrus pectin, $n = 48$; PGA, $n = 62$; known sugars, $n = 62$), (ii) enzyme type (Glucanex, $n = 70$; Viscozyme, $n = 80$; known sugars, $n = 62$), and (iii) sample and enzyme type (Glucanex-apple-POS, $n = 18$; Viscozyme-apple-POS, $n = 22$; Glucanex-citrus-POS, $n = 12$; Viscozyme-citrus-POS, $n = 36$; Glucanex-PGA-POS, $n =$

40; Viscozyme-PGA-POS, n = 22; known sugars, n = 62). All models were validated and tested on new samples.

When interpreting these models, it should be considered that RF, BLR, and MLP are computed in a different way, leading to different results. These models may show a better/ worse performance depending on the spectral data and its applications. RF builds multiple decision trees, outputting the different classes (i.e., enzyme used, pectin source), and each node is split using the best among a subset of predictors (i.e., *m*/*z* ions) randomly chosen. Then, RF averages the results from each tree to get a more accurate and stable prediction. BLR is considered as an ensemble method that uses a weighted average of predictions of individual classifiers. The term "boosting" refers to using a set of weak models, each slightly different, to build a strong model. In BLR, the base classifier (weak model) is logistic regression. The term iterations specify the maximal number of models to be fitted. On the other hand, MLP is the most common kind of ANN, one of the most popular families of machine learning models that allow one to model complex and highly nonlinear processes. MLP is formed by an input layer (i.e., *m*/*z* ions), an output layer (i.e., enzyme used, pectin source), and several neurons or nodes organized in hidden layers, where each neuron in a layer is connected with each neuron in the next layer through a weighted connection. The activation function is a transformation applied to the input spectra to determine whether the information that the neuron is receiving is relevant or not.

The training, cross-validation, and test rates for the sample type classification study were, respectively: RF, 100%, 98.6%, and 100%; BLR, 100%, 96.5%, and 100%.

Kappa values for RF and BLR were 0.98 and 0.95, respectively. Kappa is a robust measure considering the possibility of a correct classification by chance. Taking into account that no misclassifications occurred during the test phase, model sensitivity, specificity, balanced accuracy between sensitivity and specificity, and true positive and negative rates were 100%. The most influential *m*/*z* ions in each model were determined (Tables S4.2 and S4.3).

Then, POS were classified according to their enzymatic origin with the following training, cross-validation, and test rates, respectively: RF, 100%, 98.6%, and 100%; BLR, 100%, 98.6%, and 100%.

Kappa values for RF and BLR were 0.98 for both cases. Again, no misclassifications were observed, so model sensitivity, specificity, balanced accuracy, and true positives and negatives rates were 100%. Also, the most influential *m*/*z* ions in each model were determined (Tables S4.4 and S4.5).

Finally, the predictive power of these algorithms was ensured by classifying POS according to both enzyme origin and sample type. In addition, RF and BLR performance was compared to a third model that also showed high prediction rates in this study, an artificial neural network, MLP. Training, cross-validation, and test rates were, respectively: RF, 100%, 97.3%, and 100%; BLR, 100%, 97.9%, and 100%; MLP: 100%, 88.7%, and 90.6%

Kappa values for RF, BLR, and MLP were 0.97, 0.97, and 0.86 RF and BLR showed significantly higher overall accuracy and kappa values than MLP (Figure S4.4), confirming their high prediction rates. For MLP, lower model sensitivities were obtained for Glucanex-PGA-POS and Glucanex-Apple-POS (73.3% and 85.7%, respectively) while model specificity was above 95% in all cases. The true positives rate was lower for Viscozyme-PGA-POS and known sugars (71.4% and 89.5%, respectively), and the true negatives rates were higher than 90%. Balanced accuracy ranged from 85.7% to 100% (Table S4.6). For RF and BLR, these rates were 100%. The most influential *m*/*z* ions in each model were determined (Tables S4.7, S4.8, and S4.9).

It has been suggested that RF is a more robust model compared to BLR and MLP, leading to a higher accuracy when tested on new samples (test rate). This fact was also reported in our previous work (Sabater et al., 2019a). However, model performance depends on input data, so a comparison between algorithms was needed. Moreover, these algorithms reinforce each other and are different from conventional chemometric methods. PCA is not able to properly discriminate between structures obtained from specific substrates using a specific enzyme (Figure S4.5a) showing a poor performance (low percentages of variance explained by the first components). In contrast, RF, BLR, and MLP showed high classification accuracies when tested on new samples as explained above. To illustrate this high predictive power, the model performance was compared to the one obtained in a traditional supervized method like PLS-DA (Figure S4.6). It should be considered that oligosaccharide GC-EI-MS fragmentation is a complex process that produces high dimensional data that may exhibit a significant

degree of nonlinearity. In addition, differences in GC-EI-MS spectra of similar molecules like POS structures are very subtle and may not be discriminated by conventional models. Therefore, advanced pattern recognition methods are needed.

We have demonstrated that it is possible to get highly accurate classifications of POS based on several mathematical approaches. The ability to classify POS highlights structural differences according to enzyme and pectin sources.

These models discriminated POS obtained with the same enzyme preparation and different pectin, although these hydrolysates gave very similar HPSEC-RID chromatographic profiles.

Selection of the Most Relevant Fragments Using Random Forests

Machine learning algorithms detected structural differences in POS GC-MS spectra according to their enzyme and sample types and were also able to discriminate between POS obtained from a specific substrate with a specific enzyme preparation with RF being a highly accurate method. To interpret these results, RF built was also used for selecting the most relevant *m*/*z* fragments in these classifications via backward feature elimination. This method is able to determine the variables that had the strongest impact on the classification and has been used to select a set of lipids as nutritional biomarkers in infant metabolism (Acharjee et al., 2017).

Therefore, 12 fragments explained differences according to the sample type (*m*/*z* 121, 370, 514, 603, 616, 629, 646, 657, 699, 703, 733, and 747); 30 fragments explained differences according to the enzyme type (*m*/*z* 123, 135, 151, 235, 238, 252, 279, 294, 309, 310, 442, 458, 513, 516, 528, 546, 603, 604, 613, 616, 617, 630, 660, 704, 719, 731, 732, 733, 734, and 749), and 11 fragments explained differences according to both pectin and enzyme types (*m*/*z* 107, 121, 223, 252, 296, 482, 617, 646, 719, 720, and 734). Some of these fragments were influential in more than one classification, so a total of 44 *m*/*z* different fragments were considered.

Probable Structures of Relevant Fragments and Chemical Origin

Probable chemical structures of selected fragments have been suggested using CFM-ID code (Allen et al., 2016). Proposed structures are shown in Tables R4.3 and S4.10. For this purpose, a POS *in silico* fragmentation library was built by calculating complete feasible fragments for candidate molecules (i.e., pectic monosaccharides, POS-DP2, and DP3), systematically breaking bonds within the molecule (Allen et al., 2016; Allen et al., 2015). For monosaccharides, pectic neutral sugars (arabinose, rhamnose, xylose, and galactose) and GalA (methyl- or acetyl-esterified or not) were considered. The POS-DP2 included, taking into account more possible structures present in HG and RG-I, were Di-GalA (methyl-esterified or not), xylose-α(1,3)-GalA (methyl- or acetyl-esterified or not), GalA-α(1,2)-rhamnose (acetylated or not), rhamnose-α(1,4)-GalA (acetylated or not), galactose-α(1,4)-rhamnose, and arabinose-α(1,4)-rhamnose (Atmodjo et al., 2013; Caffall & Mohnen, 2009). A Tri-POS library was generated, considering combinations of all POS-DP2 structures and pectic monomers, following the main criteria reported by those authors. The criteria were established following a previous work (Sabater et al., 2019a). Then, we selected a chemical structure for each *m*/*z* ion by looking for these specific ruptures in our library. Spectral fragments, *m*/*z*, with statistically significant differences ($p < 0.05$) are reflected below.

Sample Type Study. Several *m*/*z* ions were more abundant in known compounds (pectic sugars and Di- and TriGalA). These ions include *m*/*z* 121, 370, 514, 603, 616, 646, 657, 699, 703, 733, and 747. All these fragments could be derived from GalA-α(1,4)-GalA oligomers. In addition, *m*/*z* 370 formed from GalA-containing POS was high in PGA-POS. On the other hand, the abundance of *m*/*z* 121, derived from all POS containing GalA-α(1,4)-GalA, was high in apple-POS while *m*/*z* 629 might be originated from GalA-α(1,2)-rhamnose, rhamnose-α(1,4)-GalA, and xylose-α(1,3)-GalA, and it was lower in apple-POS.

Enzyme Type Study. Like the previous study, ions *m*/*z* 294, 309, 310, 442, 458, 513, 516, 528, 546, 603, 604, 613, 616, 617, 630, 660, 704, 719, 731, 732, 733, and 734 showed higher abundances in known sugars than in unknown POS. These ions could be derived from GalA-containing POS. From them, ions *m*/*z* 309, 458, 513, 528, and 546 were previously reported by Sabater et al. (2019a) showing high abundances in known

pectic sugars too. Other ions (*m*/*z* 123, 135, 151, 235, 238, 252, and 279) were more abundant in Viscozyme-POS; from them, nonspecific ions *m*/*z* 135, 151, and 252 could be formed from all pectic sugars and POS ruptures while *m*/*z* 123 could be derived from POS containing methyl-esterified GalA. Other more specific ions are *m*/*z* 235 from POS containing neutral sugars (galactose-α(1,4)-rhamnose, arabinose-α(1,4)-rhamnose, GalA-α(1,2)-rhamnose, rhamnose-α(1,4)-GalA, and xylose-α(1,3)-GalA) and *m*/*z* 238 from acetylated GalA-α(1,2)-rhamnose and xylose-α(1,3)-GalA (acetylated), and *m*/*z* 279 might be specifically derived from acetylated GalA-α(1,2)-rhamnose. Finally, ion *m*/*z* 749, formed from GalA-containing POS, was high in Glucanex-POS.

Pectin and Enzyme Type Study. Ions *m*/*z* 617, 646, 719, 720, and 734 were high in known pectic sugars, and most of them were also characteristic for these compounds in the other comparative studies. In contrast, *m*/*z* 107, 121, 223, 252, 296, and 482 were high in Viscozyme-apple-POS; nonspecific ions *m*/*z* 107, 121, 223, and 252 are derived from GalA-containing POS while *m*/*z* 296 might be specifically originated from xylose-α(1,3)-GalA (methyl-esterified or not) rupture as well as from its two monomers, and *m*/*z* 482 was derived from GalA-α(1,2)-rhamnose (acetylated) and xylose-α(1,3)-GalA (acetylated), showing also high abundances for POS obtained with Cellulase from *A. niger* in our previous work (Sabater et al., 2019a).

These results highlight structural differences between POS and indicate that it is possible to classify the oligosaccharides formed according to the enzyme used for their obtainment and pectin source. These models could be applied on new reaction mixtures with novel oligosaccharides to determine the pectin used as raw material or study the similarities of POS obtained with different commercial enzyme preparations considering the high prediction rates on new samples.

Table R4.3. Possible chemical structures of some of the most influential *m/z* ions given by machine learning algorithms, determined by competitive fragment modelling (CFM-ID), selected for the twelve rules. They correspond to TMSO fragments from pectic di- or trisaccharides present in enzymatic hydrolysates of citrus and apple pectin and PGA.

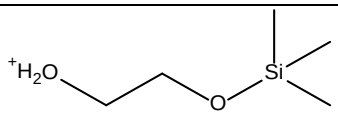
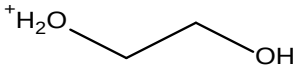
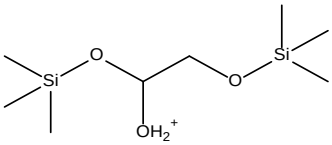
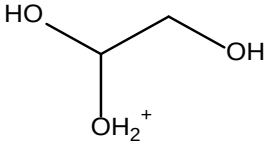
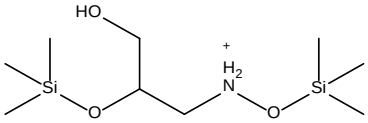
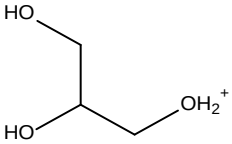
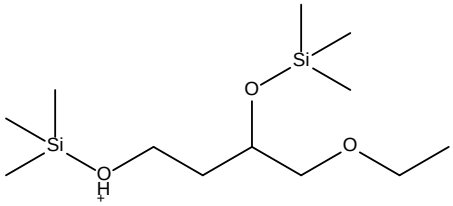
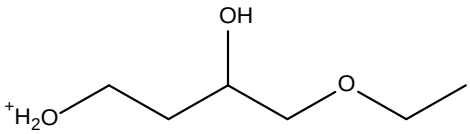
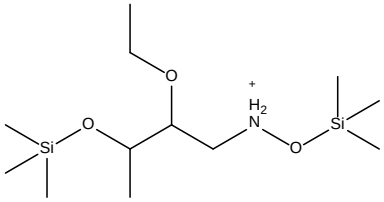
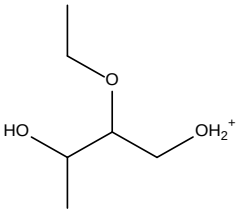
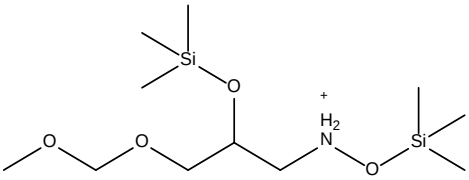
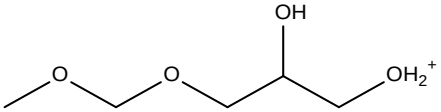
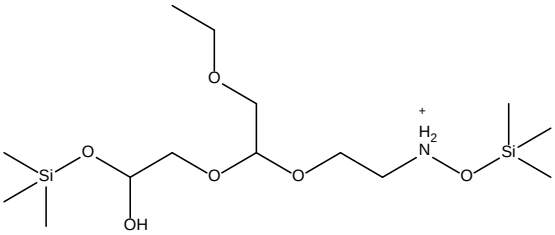
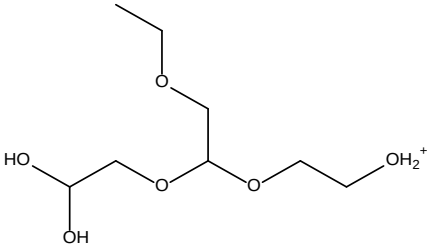
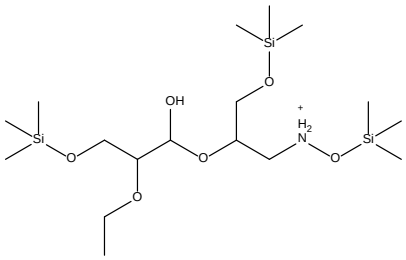
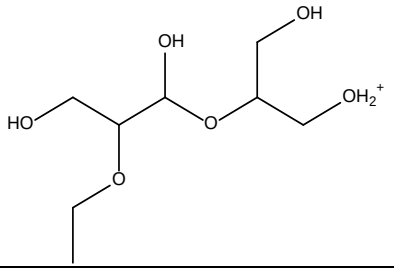
m/z	Derivatized structure*	Original molecule fragment
135		
223		
252		
279		
294		
296		
370		
442		

Table R4.3. Cont.

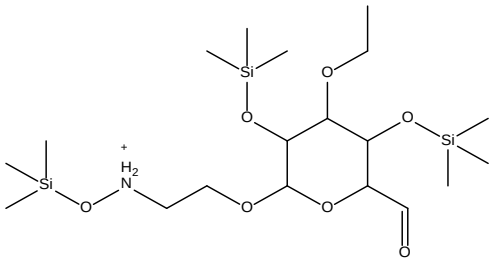
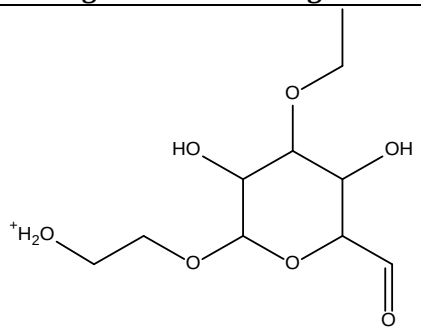
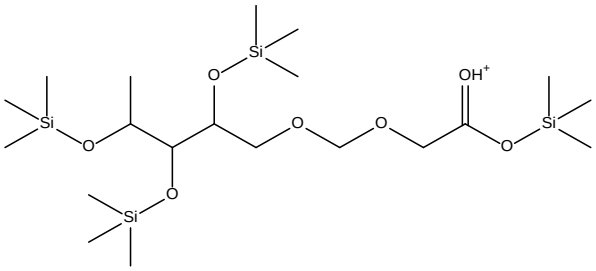
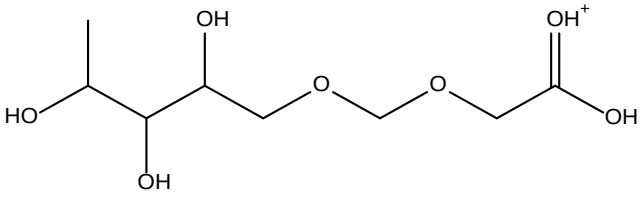
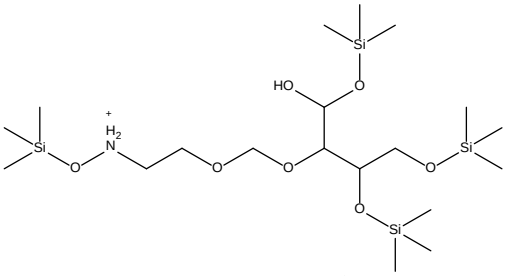
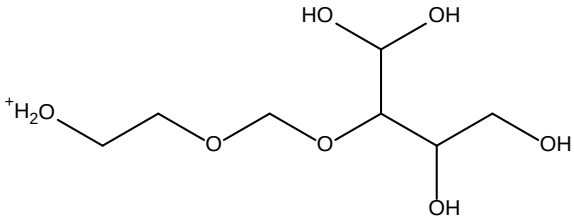
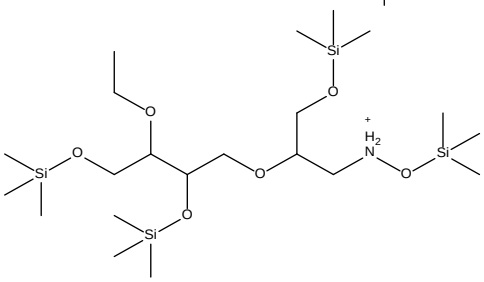
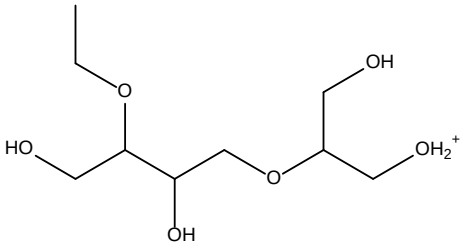
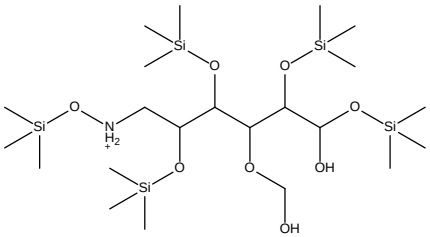
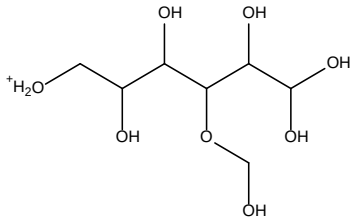
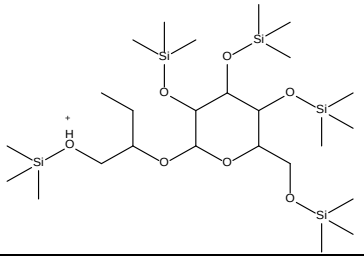
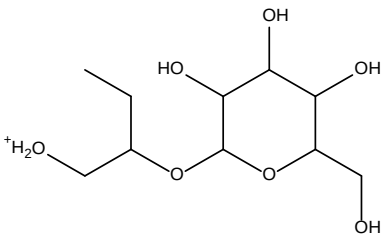
m/z	Derivatized structure*	Original molecule fragment
482		
513		
516		
528		
604		
613		

Table R4.3. Cont.

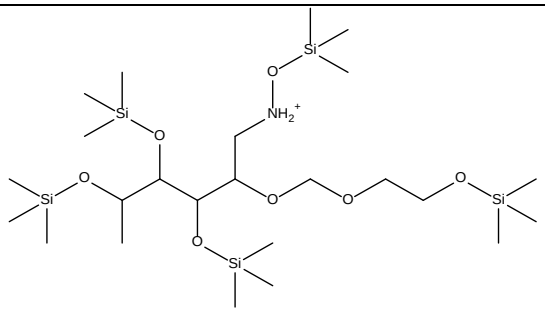
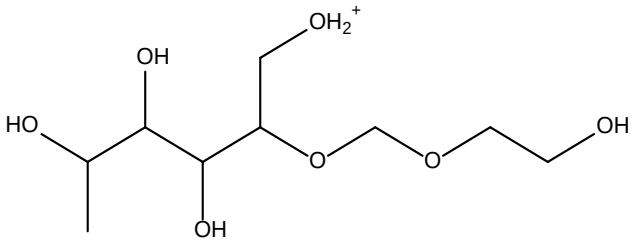
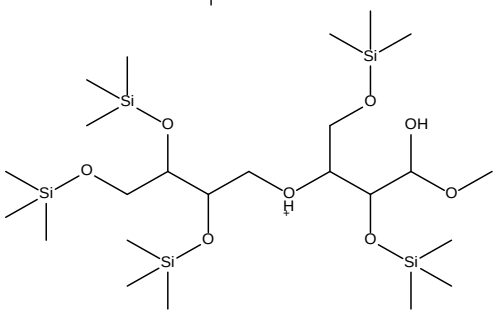
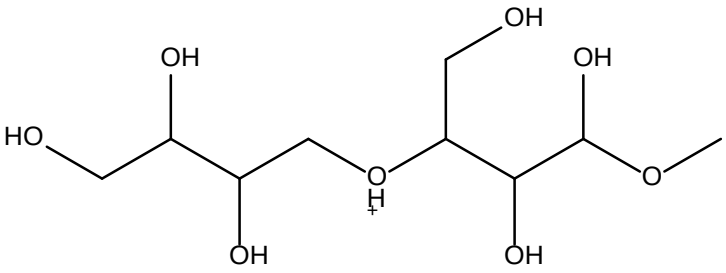
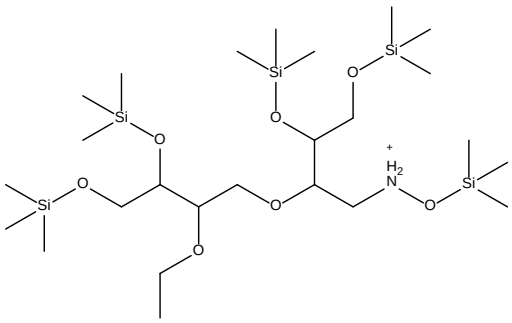
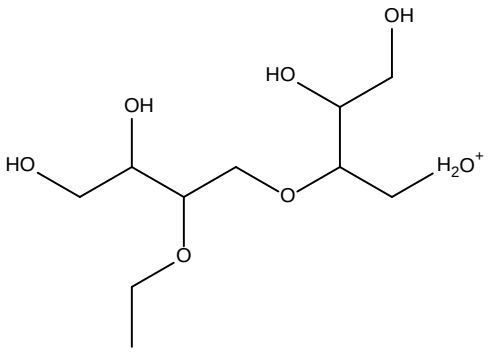
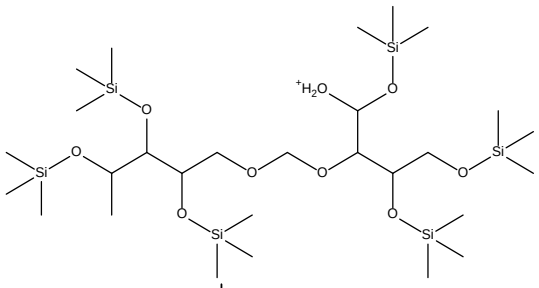
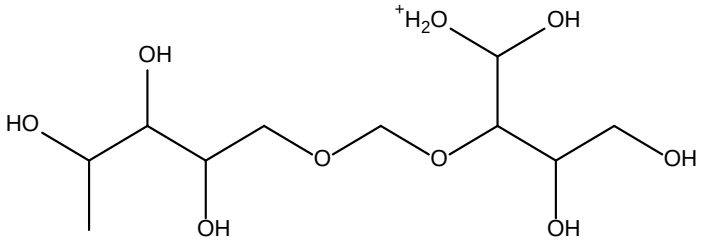
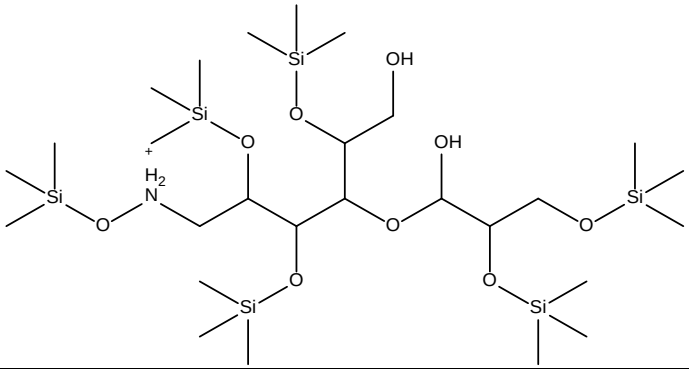
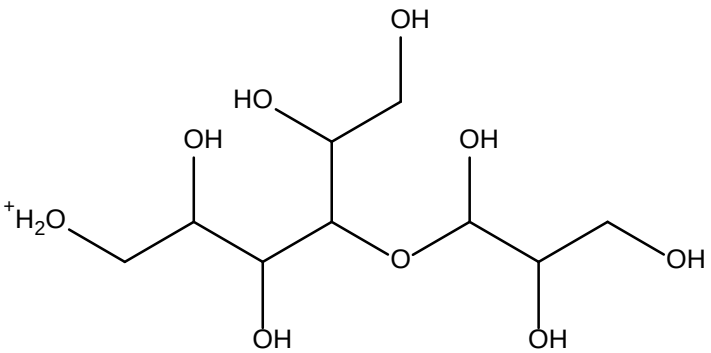
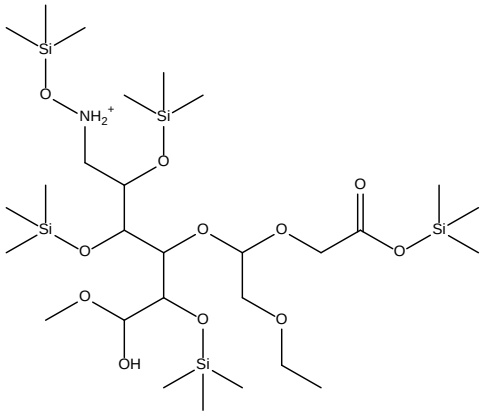
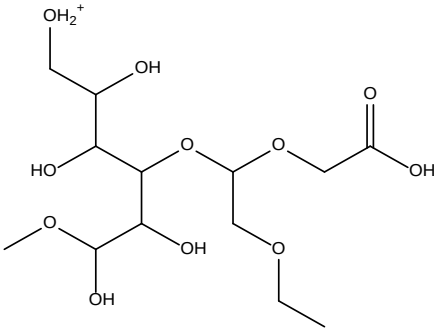
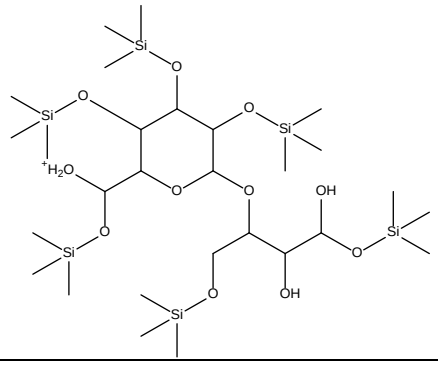
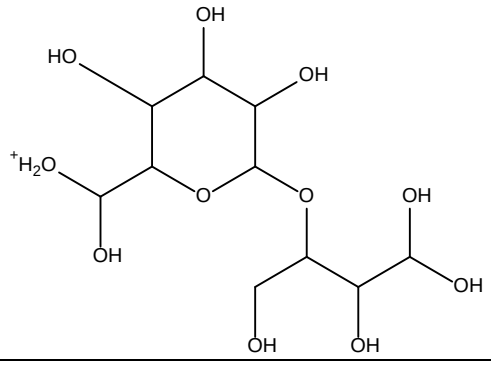
m/z	Derivatized structure*	Original molecule fragment
616		
617		
630		
719		
720		

Table R4.3. Cont.

m/z	Derivatized structure*	Original molecule fragment
734		
749		

*These structures correspond to derivatized TMSO fragments.

Association Rules between Relevant Fragments

To ensure that these models extract the most relevant structural information from GC-MS spectra, selected ions in RF studies (Table R4.3) were used to establish 12 association rules. When compared to simpler models, RF and other rule-based models provide additional information apart from variable importance analysis. RF generates classification rules based on *m*/*z* ion abundances learned directly from the input data (POS GC-EI-MS spectra) (Figure S4.7). These rules indicate possible relationships between ions originated from specific POS structures, and possible parent molecules, present in each hydrolysate, could be suggested. On the other hand, no clear relationships between specific ions related to specific hydrolysates can be proposed from the loadings plot of simple models like PCA (Figure S4.5b). However, the RF model consists of 1000 different trees, and relevant rules between *m*/*z* ions should be extracted to deepen model interpretation. To summarize the most important relationships

between POS *m/z* ions, association rules were calculated. Association rule modeling involves fewer rules and contains contextual information within the ruleset, so they can be easily interpreted. Moreover, the calculated association rules showed confidence values of 1.0; this fact means that ions described in each rule are formed together in 100% of cases, so these rules are highly reproducible and could be generalized and applied on similar oligosaccharide structures or POS from other sources. These rules (Figure S4.8) indicate which *m/z* ions originated from the same POS structures, considering both pectin and enzyme types. The first 6 rules were more characteristic of known pectic sugars.

Rule 1: *m/z* 135, 223, and 516 in known sugars. These three ions are formed from the rupture of GalA-α(1,4)-GalA units.

Rule 2: *m/z* 252, 294, 370, and 516 in known sugars. Similar to rule 1, all these ions correspond to different ruptures of GalA dimers and could be derived from *m/z* 516.

Rule 3: *m/z* 296, 370, and 613 in known sugars (probably DiGalA and TriGalA). Ion *m/z* 296 is derived from xylose or GalA ruptures and dimers of these two sugars and is originated with *m/z* 370 and 613, formed from GalA dimers or trimers bonded by α(1,4) glycosidic linkages.

Rule 4: *m/z* 296, 370, and 513 in known sugars.

Rule 5: *m/z* 296 and 734 in known sugars.

Rule 6: *m/z* 296 and 720 in known sugars.

As previously commented, *m/z* 296 corresponds to xylose or GalA ruptures while the other ions in rules 4, 5, and 6 are formed from GalA dimers or trimers. On the other hand, the rest of the rules were more characteristic of unknown POS found in hydrolysates.

Rule 7: *m/z* 279 and 528 in Glucanex-apple-POS. Ion *m/z* 279 derives from acetylated GalA-α(1,2)-rhamnose while *m/z* 528 is formed from GalA dimers. These results could indicate different ruptures of oligosaccharides containing one/two units of GalA bonded to one rhamnose molecule.

Rule 8: *m*/*z* 296, 516, and 719 in Glucanex-PGA-POS. These ions are formed from GalA chains and may also be from POS containing one/two units of GalA bonded to one xylose molecule.

Rule 9: *m*/*z* 616 and 749 in Viscozyme-citrus-POS.

Rule 10: *m*/*z* 604 and 749 in Viscozyme-apple-POS.

Rule 11: *m*/*z* 442 and 617 in Viscozyme-PGA-POS.

All ions from rules 9, 10, and 11 are formed from GalA-α(1,4)-GalA units indicating HG ruptures by Viscozyme preparation.

Rule 12: *m*/*z* 482 and 630 in Viscozyme-apple-POS. Ion *m*/ *z* 482 is derived from GalA-α(1,2)-rhamnose (acetylated) or xylose-α(1,3)-GalA (acetylated) while *m*/*z* 630 is formed from GalA chains, indicating the rupture of larger structures containing several units of GalA with rhamnose or xylose residues.

As indicated above, association rules determine which ions are formed together, and possible parent molecules may be suggested (Figure S4.8). Considering all ions associated by each rule are originated together, these ions share common regions in their structure and derive from larger oligosaccharide structures (Figure S4.9). To indicate potential parent molecules, it should be noted that these ions derive from specific POS according to our *in silico fragmentation* library. Ions from rules 1−6 derive from galacturonic acid (GalA) oligomers and are associated, indicating the presence of DiGalA and TriGalA. Fragments from rule 7 are formed from acetylated GalA-α(1,2)-rhamnose and GalA dimers, indicating the presence of acetylated DiGalA-α(1,2)-rhamnose. Ions from rule 8 derive from GalA oligomers and xylose-α(1,3)-GalA units, so xylose-α(1,3)-DiGalA is present in these hydrolysates. Fragments from rules 9−11 are relevant in unknown POS (not DiGal nor TriGalA) and derive from GalA-α(1,4)-GalA units, so the presence of trisaccharides consisting of GalA-α(1,4)-GalA attached to neutral sugars can be suggested. Finally, ions from rule 12 are formed from GalA-α(1,4)-GalA, GalA-α(1,2)-rhamnose (acetylated), or xylose-α(1,3)-GalA (acetylated) units, indicating the presence of acetylated DiGalA-α(1,2)-rhamnose or acetylated xylose-α(1,3)-DiGalA. It should be noted that these POS structures could not have been elucidated on the basis of variable importance analysis of traditional methods. Statistical

differences among relevant ions and association analysis give complementary information. It can be suggested that both Viscozyme and Glucanex produce POS containing several units of GalA and one unit of rhamnose or xylose. To verify the applicability of these rules, GC-MS spectra were classified according to both, pectin and enzyme types, in a similar way to the third machine learning study (iii) but considering only these rules (not *m*/*z* abundances). Train and test rates were 100% and 95.3%, respectively, with a kappa value of 0.94. Model sensitivity was low for Glucanex-citrus-POS (66.7%), but model specificity was above 98% for all the classes. True positives rate was low in Glucanex-apple-POS and Viscozyme-apple-POS (83.3% and 85.7%) and Viscozyme-PGA-POS (85.7%), but true negative rate was high in all cases. Balanced accuracies ranged from 83.3% to 100%, showing the lowest values for Glucanex-citrus-POS (Table S4.11).

There is a wide variety of novel pectin sources and enzymes that can be used for POS obtainment. This makes it difficult to predict POS from a specific source. However, the present study and our previous study (Sabater et al., 2019a), reporting depolymerization of low methyl-esterification pectin from artichoke (DM 19.5 ± 0.0%), allow us to tentatively predict some of the most probable POS structures that may be obtained from HG type polymer (PGA) as well as high or low methyl-esterification pectin using different enzymes. These structures are reported in Table R4.4. Interestingly, production of POS containing rhamnose and GalA dimers was correlated to the pectin content of these two sugars as well as the DM of pectin. Rhamnose-α(1,4)-GalA (acetylated or not) was produced only from low methyl-esterification pectin with high neutral sugar content, while POS containing units of GalA-α(1,2)-rhamnose were also produced from apple and citrus pectin when Viscozyme (an endo-PG) was used. Similarly, production of POS containing xylose and GalA was correlated to the DM. Xylose-α(1,3)-GalA containing POS were produced mainly from PGA (due probably to the presence of xylogalacturonan) as well as apple pectin and artichoke pectin. Arabinose containing POS are mainly produced from low methyl-esterification pectin with high neutral sugar content using only a xylanase that disrupts the RG-I branch.

Table R4.4. Most probable units found in pectic oligosaccharides (POS) structures produced from polygalacturonic acid (PGA), high methyl-esterification pectins (citrus and apple) and low methyl-esterification pectin (artichoke) by enzymatic hydrolysis using different enzyme preparations. Xyl: xylose, Ara: arabinose, Rha: rhamnose, GalA: galacturonic acid, DM: degree of methyl-esterification, PG: polygalacturonase, exo-Gal: exo-β-D-galactofuranosidase, CAN: Cellulase from *Aspergillus niger*, PUO: Pectinex Ultra Olio.

			POS containing GalA and Rha units			
Substrate	**Enzyme**	**Article**	**GalA-α(1,2)-Rha**	**GalA-α(1,2)-Rh (acetyl-esterified)**	**Rha-α(1,4)-GalA**	**Rha-α(1,4)-GalA (acetyl-esterified)**
PGA	Viscozyme (endo-PG)	Present study	Yes	Yes	-	-
PGA	Glucanex (exo-Gal)	Present study	-	-	-	-
Citrus	Viscozyme (endo-PG)	Present study	Yes	Yes	-	-
Citrus	Glucanex (exo-Gal)	Present study	-	-	-	-
Apple	Viscozyme (endo-PG)	Present study	-	Yes	-	-
Apple	Glucanex (exo-Gal)	Present study	-	Yes	-	-
Artichoke	Glucanex (exo-Gal)	Sabater et al. (2019a)	-	Yes	-	Yes
Artichoke	Pentopan (1,4-endoxylanase)	Sabater et al. (2019a)	Yes	Yes	-	Yes
Artichoke	CAN (Cellulase)	Sabater et al. (2019a)	-	Yes	-	-
Artichoke	PUO (Pectin-lyase)	Sabater et al. (2019a)	Yes	Yes	Yes	Yes

Table R4.4. Cont.

			POS containing GalA and Xy unitsl			POS containing Ara and Rha units
Substrate	**Enzyme**	**Article**	**Xyl-α(1,3)-GalA**	**Xyl-α(1,3)-GalA (methyl-esterified)**	**Xyl-α(1,3)-GalA (acetyl-esterified)**	**Ara-α(1,4)-Rha**
PGA	Viscozyme (endo-PG)	Present study	-	-	Yes	-
PGA	Glucanex (exo-Gal)	Present study	Yes	Yes	Yes	-
Citrus	Viscozyme (endo-PG)	Present study	-	-	Yes	-
Citrus	Glucanex (exo-Gal)	Present study	-	-	-	-
Apple	Viscozyme (endo-PG)	Present study	Yes	Yes	Yes	-
Apple	Glucanex (exo-Gal)	Present study	-	-	-	-
Artichoke	Glucanex (exo-Gal)	Sabater et al. (2019a)	-	-	Yes	-
Artichoke	Pentopan (1,4-endoxylanase)	Sabater et al. (2019a)	Yes	Yes	Yes	Yes
Artichoke	CAN (Cellulase)	Sabater et al. (2019a)	-	-	Yes	-
Artichoke	PUO (Pectin-lyase)	Sabater et al. (2019a)	Yes	Yes	Yes	-

CONCLUSIONS

It was observed that the same enzyme preparation applied on different pectin sources gave rise to hydrolysates with similar chromatographic profiles using traditional techniques. In contrast, these two enzymatic preparations produced POS with different profiles regardless on pectin source. Considering the HPSEC-RID results, Glucanex produced POS with a higher M_w (around 7 kDa) than Viscozyme, (around 2.4 kDa). Moreover, when POS of low M_w were analyzed by HPAEC-PAD and GC-MS, a higher amount of Di-GalA and Tri-GalA was formed with Viscozyme, showing a different profile of released neutral and acid carbohydrates. In regard to the GC-MS characterization of the POS structures, this work should be considered as a first approach to a comprehensive study. Enzymatic depolymerization of two common highly methyl-esterified pectins (from citrus and apple) as well as PGA, used as HG standard, employing two enzymes with very different activities was studied. Thus, POS characterization using a relatively simple technique like GC-EI-MS allows one to verify structural similarities and differences in POS obtained from the most widely used pectin sources. Mass spectra of di- and trisaccharides derived from GalA have been obtained from apple and citrus pectin and PGA for the first time. In addition, characteristic GC-EI-MS patterns were attributed to the presence of specific units in POS structures. However, only small oligosaccharide molecules are detected by GC-MS. Dimers and trimers described in this study are also present in larger molecules produced during enzymatic hydrolysis. To characterize larger structures, more studies dealing with more advanced MS-based techniques are needed.

The results presented here improve existing computational methods by extracting more information from oligosaccharide GC-EI-MS spectra. Finally, this methodology could be applied to novel reaction mixtures or oligosaccharides from unknown origin obtained with different enzymes to determine similar hydrolysis patterns and find possible structure−function relationships.

Chapter III:

Practical applications of artichoke pectin and pectic oligosaccharides

Article V

6.3. Chapter III: Practical applications of artichoke pectin and pectic oligosaccharides

6.3.1. Article V: GC-EI-MS and MALDI-TOF-MS characterization of the carbohydrate fraction of functional yogurts containing pectin and pectic oligosaccharides. Applications of structure-retention time relationships and convolutional networks

GC-EI-MS and MALDI-TOF-MS characterization of the carbohydrate fraction of functional yogurts containing pectin and pectic oligosaccharides. Applications of structure-retention time relationships and convolutional networks

Carlos Sabater, Celia Abad-García, Paloma Delgado-Fernández, Nieves Corzo, Antonia Montilla

Instituto de Investigación en Ciencias de la Alimentación CIAL, (CSIC-UAM) CEI (UAM + CSIC), C/Nicolás Cabrera, 9, E-28049 Madrid, Spain

A B S T R A C T

The carbohydrate fraction of functional yogurts supplemented with citrus and artichoke pectin and their pectic oligosaccharides (POS) has been characterized, including carbohydrates form raw material (milk, pectin and POS) and formed during yogurt manufacture (GOS and POS). Lactose and lactic and acetic acid contents were in

the range 3.5-4.0; 0.85-1.16 and 0.04-0.07 g 100 g^{-1}, respectively. GC-EI-MS spectra of yogurt carbohydrates were classified using machine learning and structure-relative retention time relationships were calculated to determine the abundance of specific fragments on larger oligosaccharide structures. All information generated was correlated using a convolutional neural network that established characteristic patterns in the complete carbohydrate fraction of each yogurt. It was found that di-, tri- and tetra-POS formed by rhamnose and xylose attached to acetylated-galacturonic acid were released during fermentation in yogurts with artichoke POS. Structures elucidated by these algorithms were confirmed by MALDI-TOF-MS. These models allow determining structural differences among novel oligosaccharides present in food matrices.

Keywords: Functional yogurt, Artichoke pectin, Citrus pectin, Pectic oligosaccharides, *in silico* fragmentation, Convolutional neural network

INTRODUCTION

Yogurt is one of the most popular fermented milk consumed worldwide which improves equilibrium of microbiota besides being a good vehicle to delivery functional ingredients (Demirkol & Tarakci, 2018). Among them, prebiotic carbohydrates have been incorporated in yogurt and kefir (Delgado-Fernández et al., 2019a; Delgado-Fernández et al., 2019b). Pectins are considered as emergent prebiotics and present interesting functional activities which allow their use as food additive. Most works are focused on the study of technological properties in pectin-supplemented yogurts. In fact, when yogurts were supplemented with citrus pectin an enhancement of rheological quality has been perceived, favouring proliferation of *Streptococcus thermophilus* and *Lactobacillus bulgaricus* during the period of fermentation (Arioui et al., 2017). While high-methyl-esterified pectin produces significantly higher thickness than other hydrocolloids (Gallardo-Escamillaa et al., 2007), applications of low-methyl-esterified pectin have been also studied, increasing oral consistency in goat milk yogurts (Bruzantin et al., 2016). However, few research addresses the characterization of the

carbohydrate fraction and their possible changes during fermentation and storage of yogurts supplemented with pectin, on the other hand, yogurts with POS have not been studied yet.

In general, pectin and POS present very complex structures and to characterize them several chromatographic techniques (Sabater et al., 2019a; Muñoz-Almagro et al., 2017) and advanced data analysis strategies have been used (Sabater et al., 2019a). However, mass spectrometry (MS) hyphenated techniques are the preferred. To interpret GC-EI-MS fragmentation patterns, *in silico* fragmentation methods have been proposed (Allen et al., 2016) which can be used in combination with machine learning algorithms to determine POS substructures obtained with different enzyme preparations and pectin sources (Sabater et al., 2019a; Sabater et al., 2019b). These computational tools have been also applied to characterise original pectin structures extracted by different methods (Sabater et al., 2020). Other advanced MS-based techniques include MALDI-TOF-MS, which has been used to elucidate larger structures of galactan oligomers (Zheng et al., 2018).

To gain a deeper knowledge of structure of molecules under study, quantitative structure-retention time relationships may be determined. This methodology has been applied on GC×GC/QMS analysis of food samples (Rojas et al., 2019) and LC-MS based on untargeted analysis (Yu et al., 2019). Another powerful model for MS sample fingerprinting are convolutional neural networks (CNN), an image recognition model recently applied in automated targeted analysis of raw GC-EI-MS data to classify volatile organic compounds (Skarysz et al., 2018) and more advanced MS imaging techniques for tumour classification (Behrmann et al., 2017).

Therefore the aim of this work was to study the behaviour of carbohydrates during elaboration and storage of yogurts supplemented with citrus or artichoke pectin and POS. To this end, a comprehensive characterization of the carbohydrate fraction of each type of functional yogurt has been carried out using GC-EI-MS and MALDI-TOF-MS as complementary techniques. This characterization study consisted in the following steps: i) GC-EI-MS spectra classification of individual carbohydrates using machine learning algorithms to determine characteristic fragments, ii) Quantitative structure-retention time relationships for all yogurt sugars to determine the abundance of these

fragments on larger oligosaccharide structures, some of them released during fermentation, iii) CNN modelling to correlate all information generated and to compare the complete carbohydrate fraction of functional yogurts, iv) MALDI-TOF-MS analysis to confirm the structures elucidated by these models.

MATERIALS AND METHODS

Standards and reagents

Analytical reference substances such as *myo*-inositol, L-arabinose, D-galactose, D-glucose, galacturonic acid (GalA), digalacturonic acid (Di-GalA), kestose, nystose and phenyl-β-glucoside were purchased from Sigma Aldrich (Steinheim, Germany). Trigalacturonic acid (Tri-GalA) was from Carbosynth (Compton, UK). Commercial citrus pectin was acquired from Acofarma (Terrassa, Barcelona). Artichoke pectin was previously extracted in our laboratory using Celluclast®1.5L (Sabater et al., 2018a). Citrus and artichoke POS mixtures were obtained by enzymatic depolymerization of citrus/artichoke pectin using Cellulase from *Aspergillus niger* (Sigma Aldrich), (Sabater et al., 2019a). Characterization of pectin/POS functional ingredients is shown in Supplementary Material Table S5.1. Commercial starter culture (YoFlex® Advance 2.0) containing a mixture of *S. thermophilus* and *L. delbrueckii ssp. bulgaricus* was from Chr. Hansen (Hørsholm, Denmark).

Yogurt elaboration

Yogurt starter culture was prepared following the method of Delgado-Fernández et al. (2019a). Each batch of control (set-style yogurt) and functional yogurts supplemented with 1% (w/v) of citrus pectin (CP); artichoke pectin (AP); citrus POS (CPOS); or artichoke POS (APOS) was prepared in duplicate. After, they were divided into seven portions of 50 mL and incubated at 43 °C during 5 h. Samples were taken for analysis every hour during fermentation process. After incubation, yogurts were cooled to room temperature and stored at 4 °C for 1 week.

Analytical determinations

Yogurt pH, viable bacterial count and analysis of organic acids (HPLC-UV) were carried out as previously described Delgado-Fernández et al. (2019a). Before chromatographic analysis, fat and protein interfering materials were removed by precipitation using Carrez reagents. Carbohydrates were analysed by GC-EI-MS as trimethyl silyl oximes (TMSO) following the method of Sabater et al. (2019a). Two different analysis methods were followed to determine lactose (Method I) and minor yogurt carbohydrates (Method II). In Method I samples were injected in split mode (100:1). The oven temperature was initially 150 ºC, then increased at a rate of 10 ºC min^{-1} to 220 ºC, then increased at a rate of 3 ºC min^{-1} to 310 ºC and held for 5 min. In Method II samples were injected in split mode (5:1). The oven temperature was initially 150 ºC and held for 17 min, then increased at a rate of 1 ºC min^{-1} to 165 ºC and then increased at a rate of 3 ºC min^{-1} to 310 ºC and held for 30 min. In Method II, signal recording was interrupted between 61-63 min, retention time at which lactose was eluted. To corroborate GC-EI-MS results, POS present in yogurt samples were also characterized by MALDI-TOF-MS on a Voyager DE-PRO mass spectrometer (Applied Biosystems, Foster City, CA) at the Servicio Interdepartamental de Investigación (SIdI-UAM) of Madrid. Mass spectra were obtained in positive ion mode over the *m/z* range 300-1500.

GC-EI-MS and MALDI-TOF-MS structure elucidation of yogurt carbohydrates

To determine structural differences in the carbohydrate fraction of each kind of yogurt (set-style, CP, AP, CPOS and APOS) a study of spectra obtained by GC-EI-MS and MALDI-TOF-MS was carried out. Our data analysis strategy consisted in the following steps:

i) GC-EI-MS spectra classification using machine learning. Full spectra (n=521) of known sugars found in yogurts as well as GOS and unidentified POS were selected (Supplementary Material Table S5.2). To extract chemically relevant information from GC-EI-MS spectra, 110 fragments in the range of *m/z* 150-543 whose abundances were statistically different among groups ($p < 0.05$) were chosen. These fragments might correspond to known GOS and POS ruptures assessed by competitive fragmentation

modelling (CFM-ID) code (Sabater et al. 2019a; Sabater et al. 2019b; Allen et al., 2016). Discrete wavelet transform (DWT) was applied to denoise the signal (Sabater et al., 2019a). A 15 level DWT was computed and DWT coefficients under percentile 15% were removed. Then, two tree-based models, bagged classification and regression trees (BCART) and random forest (RF), were computed to classify GC-EI-MS spectra into classes shown in Supplementary Material Table S5.2. In BCART, several tree models are trained on a subset of samples and the average prediction from each model is calculated. In RF, a multitude of decision trees are constructed, outputting the different classes. Each node is split using the best among a subset of predictors randomly chosen. BCART was built with 25 bootstrap replications, while RF was built with 500 trees and 38 variables tried at each split. Chemical structures of 26 relevant fragments in the classification study were calculated (Sabater et al., 2019a; Sabater et al., 2019b)

ii) Quantitative structure-retention time relationships. Quantitative relationships between relative retention time (RRT, considering the internal standard, phenyl-β-glucoside) of yogurt carbohydrates and 26 relevant *m/z* ions (selected in the previous step) were established using BCART and RF. In addition, these algorithms were compared to a quantile regression forests (QRF), a generalization of RF models that estimates the conditional quantiles of response variables (RRT) given the *m/z* ions by building RFs. In this case, RF and QRF were built with 8 variables tried at each split.

iii) Convolutional neural network (CNN) for establishing structural patterns of carbohydrate fraction of functional yogurts. To summarize all information extracted from GC-EI-MS spectra, abundance heatmaps of selected fragments and RRT of all sugars were generated considering spectral data of the carbohydrate fraction of five different yogurts (elaborated in duplicate) at different times: initial time, end of the fermentation (5 h), and 1 week of cold storage. To accurately verify structural differences in the complete carbohydrate fraction of these yogurts (and not individual compounds studied in the two previous sections), these images were classified using a CNN to differentiate the profiles of yogurts containing: i) citrus pectin/POS (CNN1), ii) artichoke pectin/POS (CNN2), iii) specifically APOS (CNN3), from the rest of yogurts. In addition, a fourth model (CNN4) was built to discriminate between yogurt samples taken at the beginning of the fermentation (0 h) and samples taken after fermentation

and storage (5 h and 1 week). Therefore, structural profiles of oligosaccharides formed during fermentation were elucidated and corroborated mathematically. CNN architectures consisted in 2 convolution layers, 1 max pooling layer, 1 flatten layer and 1 output layer. First convolution layer captures the low-level features of the images while second layer captures the high-level features as well. The number of output filters in the convolution was 32 and 16 for the first and second convolution layer, respectively. The width and height of the 2D convolution window were 3 x 3 in these two layers. Pooling layer is used for dimensionality reduction and extracting dominant features. In this case, max pooling, which returns the maximum value from the portion of the image covered by the cluster of neurons, was applied. Flatten layer, which extracts non-linear combinations of the high-level features, had 100 units. The activation function for convolution and flatten layers was ReLu (rectified linear unit) while softmax function was employed for output layer.

iv) MALDI-TOF-MS. To confirm the structural patterns elucidated by these models from GC-EI-MS spectra of yogurt carbohydrates, molecular ions *m/z* corresponding to specific oligosaccharide chains were determined, considering an accuracy of ± 0.3 Da.

All models were trained on 70% of data and tested on 30% new samples. All statistical analyses were computed on R v3.5.0. DWT was performed using wavelets package (Aldrich, 2013). BCART was built with ipred package (Peters & Hothorn, 2018). RF classification was performed with randomForest package (Liaw & Wiener, 2002). QRF was built with `quantregForest package (`Meinshausen, 2017`).` CNN was computed using CPU-TensorFlow (Allaire & Tang, 2018) and keras (Allaire & Chollet, 2018).

RESULTS AND DISCUSSION

Set-style and functional yogurt analysis

Set-style (control) and four functional yogurts supplemented with 1% citrus and artichoke pectin (CP and AP) and POS (CPOS and APOS) were prepared. The pH of control and functional yogurts reached values in the range of 4.5-4.6 after 5 h of fermentation and a slight pH decrease was observed during 1 week of cold storage (4.2-

4.4) (Supplementary Material Figure S5.1). On the other hand, viable cell count accomplished the quality standards required for yogurts and was in the range 7.9-9.3 log_{10} cfu mL^{-1} and kept constant during the cold storage (Supplementary Material Figure S5.2). *L. bulgaricus* count in yogurts containing citrus POS was slightly higher (8.1 log_{10} cfu mL^{-1}) than in control ones (7.4 log_{10} cfu mL^{-1}) while, *S. thermophilus* count was lower in functional yogurts with citrus and artichoke POS (8.2 and 8.0 log_{10} cfu mL^{-1}) than in control (9.2 log_{10} cfu mL^{-1}). However, these differences were not statistically significant. The main carbohydrates found in all types of prepared yogurts are shown in Figure R5.1. Lactose content at the end of the storage was in the range of 3.5-4.0 g 100 g^{-1} (Supplementary Material Figure S5.3a). Although during fermentation there were significant differences between set-style and AP yogurts and the rest of functional yogurts, these differences disappeared after storage. Glucose and galactose were released as consequence of lactose hydrolysis and increased during fermentation with the exception of glucose in set-style and APOS added yogurts. Levels of both monosaccharides were very different being glucose in lower amount than galactose (Supplementary Material Figure S5.3b and S5.3c).

Both monosaccharides were present in the lowest amount in yogurts containing APOS, 0.05 and 0.50 mg 100 mg^{-1} of glucose and galactose respectively. On the other hand, glucose seems to be efficiently metabolized in the control yogurt, perhaps due to the higher content of *S. thermophilus*.

With respect to minor carbohydrates from milk (Table R5.1) *myo*-inositol was quantified (5.1-6.3 mg 100 g^{-1}) (Table R5.1). During fermentation galacto-oligosaccharides (GOS) such as allolactose (Gal-β(1,6)-Glc) and 6-galactobiose (Gal-β(1,6)-Gal) were formed in all yogurts, as a consequence of hydrolysis and transgalactosylation of lactose by β-galactosidase from lactic acid bacteria of starter culture. Allolactose and 6-galactobiose content, reached the highest values in yogurts with citrus pectin: 16.6 and 11.1 mg 100 g^{-1}, respectively. In contrast, these GOS were low in APOS-supplemented yogurts although the cause is uncertain. These GOS were the most abundant found in yogurt and other fermented milks (Martinez-Villaluenga et al., 2008; Toba et al., 1983). Other unidentified GOS formed ranging from 0.8 to 1.9 mg 100 g^{-1} were also quantified in set-style and yogurts with citrus pectin.

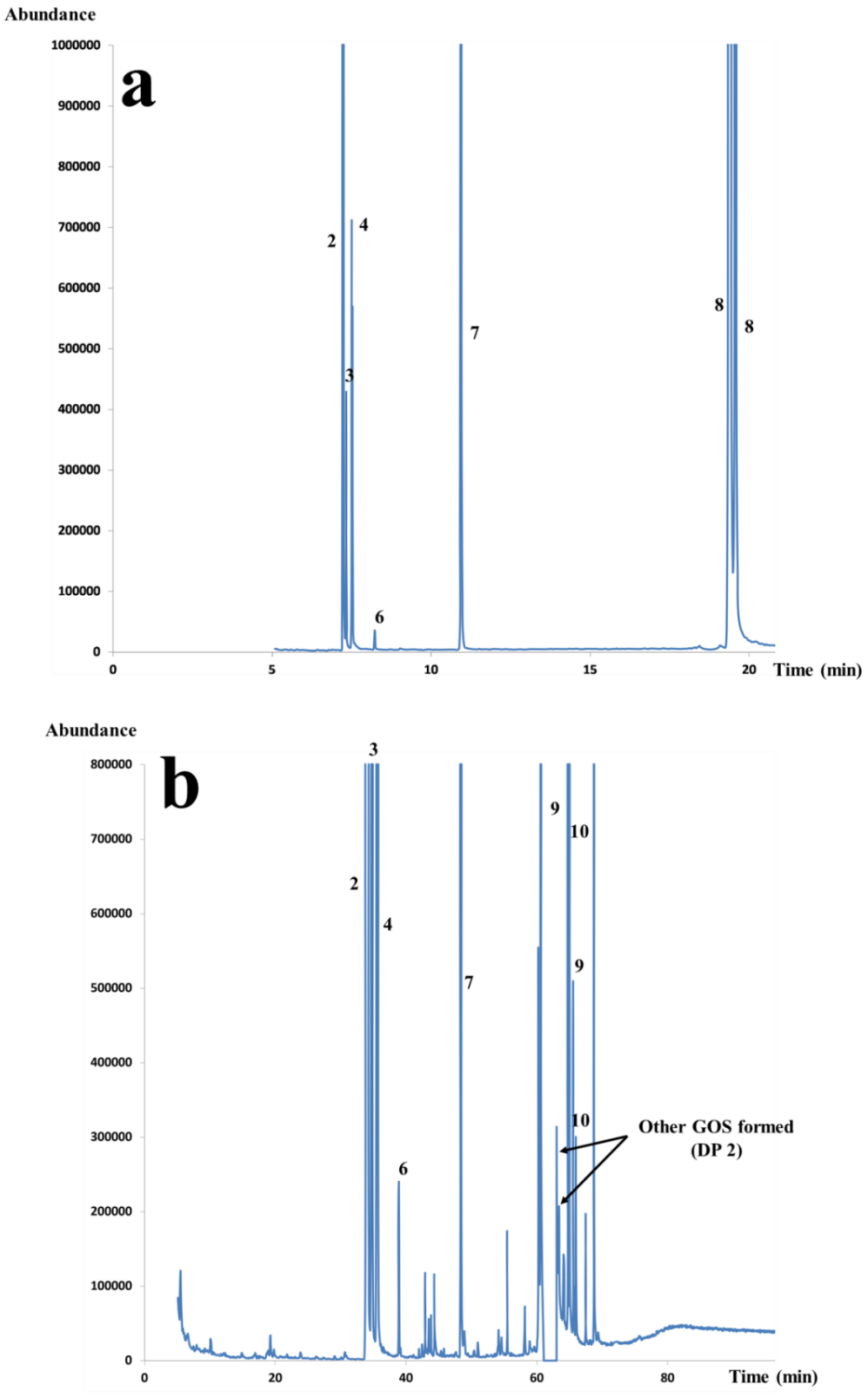


Figure R5.1. GC-MS chromatograms of yogurts, at the end of the fermentation step, following two methods for determining the main (method I, split 100:1) (**a**) and minor (method II, split 5:1) (**b**) carbohydrates of set-style yogurt; and main (**c**) and minor (**d**) carbohydrates of artichoke POS-containing yogurt. **1**: arabinose, **2**: galactose, **3**: glucose, **4**: galactose+glucose, **5**: galacturonic acid, **6**: *myo*-inositol, **7**: internal standard, **8**: lactose, **9**: alollactose, **10**: 6-galactobiose. **DP**: degree of polimerization, **APOS**: artichoke pectic oligosaccharides, **-I**: initial oligosaccharides, **-F**: formed oligosaccharides. *****POS formed during fermentation.

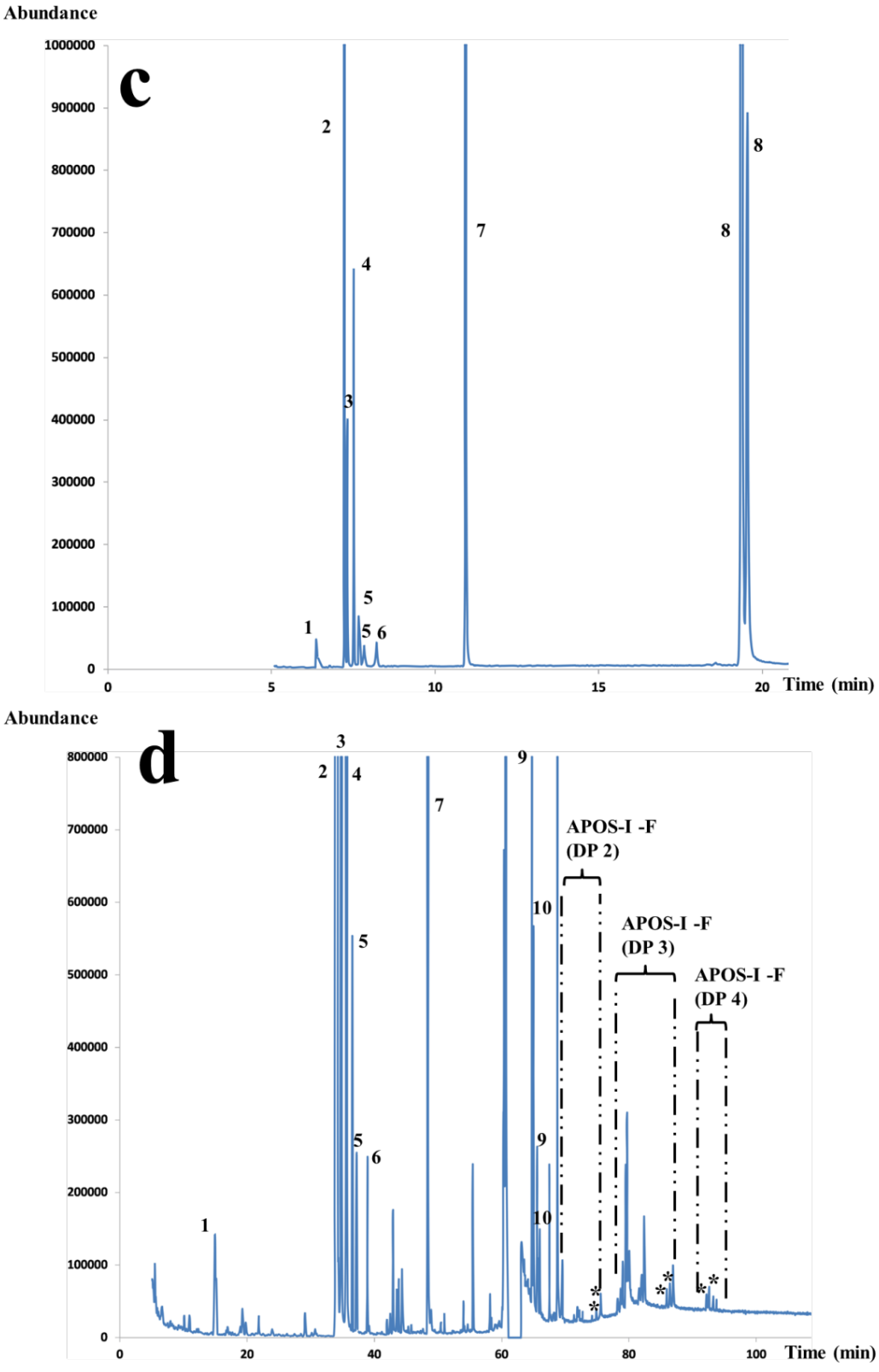


Figure R5.1. Cont.

With regard to pectic sugars, free galacturonic acid (GalA) was found in all functional yogurts (Table R5.1) as expected, the lowest amount was found in yogurts with pectins (citrus and artichoke) (10-15 mg 100 g^{-1} yogurt). In APOS y CPOS supplemented yogurts, GalA was found in high amounts 158 and 217 mg 100 g^{-1}, respectively, and these values decreased during fermentation and cold storage. Arabinose, a pectic monosaccharide, was also found in higher amount in yogurts with pectin or POS from artichoke (Table R5.1). Also, in yogurts supplemented with POS from citrus and artichoke, initial POS content ranged from 6.2 to 61.4 mg 100 g^{-1}, showing the highest values in yogurts containing these ingredient mixtures, especially APOS. In addition, formation of new POS (from di- to tetrasaccharides) in very small amounts (0.9-3.4 mg 100 g^{-1}) was detected. It should be noted that initial POS content in CPOS-yogurts did not suffer significant changes during fermentation, although the quantity in these samples was very low, while in APOS-yogurts, POS content decreased from 61.4 to 9.7 mg 100 g^{-1}. In these yogurts, a great number of different POS structures were also released during fermentation and, even though it was possible to quantify them, their quantity was minimal (Supplementary material Table S5.2).

Table R5.1. Minor carbohydrates found in set style and supplemented citrus and artichoke pectins and POS yogurts. **Ara**: arabinose, **GalA**: galacturonic acid, **AloLa**: allolactose, **Gabio**: 6-galactobiose, ***Myo*-In**: myo-inositol, **GOS**: galacto-oligosaccharides, **POS**: pectic oligosaccharides**, CP**: yogurts with citrus pectin, **AP**: yogurts with artichoke pectin, **CPOS**: yogurts with citrus POS, **APOS**: yogurts with artichoke POS, **-I**: initial oligosaccharides, **-F**: formed oligosaccharides. [a,b,c,d,e] Statistically significant differences between samples.

Yogurts	Process	Time	mg 100 g^{-1} Yogurt												
			Ara	GalA	AloLa	Gabio	*Myo*-In	GOS-F	CP-I	CP-F	AP-F	CPOS-I	CPOS-F	APOS-I	APOS-F
Set-style	Fermentation (h)	0	-	-	-	-	5.1^{a} (0.5)	-	-	-	-	-	-	-	-
		5	-	-	16.2^{a} (0.3)	9.0^{b,c} (0.1)	5.9^{a} (0.1)	0.8^{b} (0.1)	-	-	-	-	-	-	-
	Cold storage (week)	1	-	-	16.2^{a} (0.2)	9.6^{a,b,c} (0.1)	6.2^{a} (0.0)	1.0^{b} (0.3)	-	-	-	-	-	-	-
Citrus pectin	Fermentation (h)	0	0.29^{e} (0.01)	14.5^{d} (0.8)	-	-	5.2^{a} (0.2)	-	1.0^{b} (0.0)	-	-	-	-	-	-
		5	0.28^{e} (0.03)	10.6^{d} (0.4)	15.9^{a} (0.6)	8.2^{c} (0.4)	6.1^{a} (0.1)	0.9^{b} (0.0)	1.0^{b} (0.0)	0.8^{a} (0.2)	-	-	-	-	-
	Cold storage (week)	1	0.30^{e} (0.01)	10.1^{d} (0.0)	16.6^{a} (2.2)	11.1^{a} (0.4)	5.9^{a} (0.0)	1.9^{a} (0.3)	3.0^{a} (0.0)	1.0^{a} (0.3)	-	-	-	-	-
Citrus POS	Fermentation (h	0	1.04^{d} (0.08)	217.2^{a} (12.0)	-	-	5.0^{a} (0.4)	-	-	-	-	6.2^{c} (0.1)	-	-	-
		5	1.10^{d} (0.02)	159.3^{b} (5.8)	14.3^{a} (2.3)	5.1^{d} (0.5)	6.3^{a} (0.5)	-	-	-	-	8.0^{b} (0.4)	1.0^{a} (0.0)	-	-
	Cold storage (week	1	1.01^{d} (0.08)	151.1^{b} (0.1)	15.1^{a} (0.6)	9.0^{b,c} (0.5)	7.0^{a} (0.7)	-	-	-	-	9.6^{a} (0.1)	1.0^{a} (0.0)	-	-
Artichoke pectin	Fermentation (h)	0	2.08^{c} (0.09)	10.0^{d} (1.6)	-	-	5.2^{a} (0.3)	-	-	-	-	-	-	-	-
		5	2.40^{c} (0.02)	6.9^{d} (0.1)	16.5^{a} (0.6)	9.8^{a,b} (0.3)	5.8^{a} (0.3)	-	-	-	3.3^{a} (0.2)	-	-	-	-
	Cold storage (week)	1	2.41^{c} (0.02)	6.0^{d} (0.0)	17.3^{a} (0.2)	10.2^{a,b} (0.0)	6.1^{a} (0.0)	-	-	-	3.4^{a} (0.1)	-	-	-	-
Artichoke POS	Fermentation (h)	0	3.99^{b} (0.28)	158.3^{b} (24.5)	-	-	5.2^{a} (0.1)	-	-	-	-	-	-	61.4^{a} (0.1)	-
		5	4.04$_{b}$ (0.07)	103.4^{c} (2.0)	7.2^{b} (0.0)	4.3^{d} (0.2)	6.2^{a} (0.1)	-	-	-	-	-	-	33.9^{b} (1.6)	0.9^{a} (0.1)
	Cold storage (week)	1	4.55^{a} (0.05)	90.4^{c} (0.3)	7.9^{b} (0.4)	5.9^{d} (0.3)	5.9^{a} (0.4)	-	-	-	-	-	-	9.7^{c} (0.1)	0.8^{a} (0.1)

These results indicate that yogurt starter may, besides metabolizing the GalA, modify initial POS and released new POS structures in small amounts.

Regards to organic acid formation, lactic acid was the main formed, derived from metabolism of lactic acid bacteria, also acetic acid was detected (Supplementary Material Table S5.3). Their content increased with fermentation time and cold storage up to reach values in the range from 846 to 1156 and 35 to 67 mg 100 g^{-1}, respectively. Lactic acid formation was lower in yogurts with citrus POS. However, acetic acid production was high in these yogurts (57 mg 100 g^{-1}) as well as yogurts containing artichoke POS (67 mg 100 g^{-1}).

GC-EI-MS characterization of the carbohydrate fraction of set-style and functional yogurts confirmed by MALDI-TOF-MS analysis

Structural patterns in the carbohydrate fraction of functional yogurts have been established and the release of specific POS structures during yogurt manufacture was studied. For this purpose, several machine learning models were used to interpret GC-EI-MS spectra of yogurt carbohydrates and the results obtained were confirmed by MALDI-TOF-MS.

Carbohydrate spectra classification using machine learning algorithms. First, GC-MS spectra (n= 521) of monosaccharides, GOS and POS (from di- to tetrasaccharides) were collected (Supplementary Material Table S5.2). In order to study the possible structural patterns of different carbohydrates present in yogurt, the first step in our GC-MS analysis was to establish a spectra classification in 16 categories (Supplementary Material Table S5.2) considering their chemical nature:

Common carbohydrates present in milk and yogurt including standards: ***myo*-inositol** (n=31), **galactose** (n=32), **glucose** (n=32), **lactose** (n=62).

Pectic monosaccharides found in yogurts containing pectin and POS including standards: **arabinose** (n=50) and **galacturonic acid** (n=26).

Formed GOS: **allolactose** (n=40) and **6-galactobiose** (n=40). Some unidentified di-**GOS** (n=10) were formed during fermentation/storage in set-style yogurts and those containing citrus pectin.

Oligosaccharides present in citrus pectin yogurts: some POS were present at the beginning of the yogurt elaboration coming from the pectin (1 di-POS, **CP-I**; n=6); or formed (**CP-F**; n=12) during fermentation/storage (up to 2 di-POS and 2 tri-POS).

Oligosaccharides present in artichoke pectin yogurts: up to 5 different di-POS (**AP-F**; n=16) were released during fermentation/storage.

Oligosaccharides present in citrus POS yogurts: most POS were present at the beginning of the process, including 5 di-POS and 3 tri-POS (**CPOS-I**; n=48) and only 2 tri-POS were formed (**CPOS-F**; n=8) during fermentation/storage.

Oligosaccharides present in artichoke POS yogurts: this was the most interesting yogurt with regard to POS present, many of them were at the beginning of the process (**APOS-I**; n=88) corresponding at 18 initial POS (6 di-POS, 10 tri-POS and 2 tetra-POS); while up to 6 different POS peaks were formed (**APOS-F**; n=20), being 2 di-POS, 2 tri-POS and 2 tetra-POS.

To accurately classify these spectra and to extract information about its structural characteristics, two algorithms were employed: BCART and RF. The training, 10-fold cross-validation and test rates were 100, 93.4 and 96.8% for BCART; and 100, 95.3 and 98.7% for RF. Kappa values, a robust measure considering the possibility of a correct classification by chance, for BCART and RF were 0.93 and 0.95, respectively. Accuracy and kappa showed no significantly differences between models (Supplementary material Figure S5.4). When tested on new samples (Supplementary material Tables S5.4 and S5.5), these two models showed high sensitivity values (above 90%) with the exception of lactose and unidentified GOS for BCART and RF classification, respectively (86 and 75%). Therefore, these two classes showed the lowest balanced accuracy (93 and 88%). However, specificity rates were above 97% for every class in both models. It should be noted that BCART showed balanced accuracies above 98% for all different GOS and POS studied. These two models found characteristic and reproducible patterns for different types of carbohydrates (GOS and POS) and were also able to discriminate between initial oligosaccharides and those formed during fermentation and storage steps, confirming differences in their structures. Moreover, oligosaccharides were also differentiated according to the type of pectin (citrus and artichoke pectin).

To study MS patterns established by RF and BCART, the most influential ions for each classification were determined (Supplementary Material Tables S5.6 and S5.7). Then, 26 relevant fragments that explained differences in GC-MS spectra of these compounds were selected to continue our study (Supplementary Material Table S5.8). Probable chemical structures of most specific ions (TMSO derivatized structures and original carbohydrate fragments) were elucidated (Supplementary Material Table S5.8).

It should be noted that these structures contain nitrogen and trimethylsilyl groups because they correspond to TMSO fragments of carbohydrates that could be used as GC-MS marker ions. For this purpose, an *in silico* fragmentation library was built by calculating complete feasible fragments for candidate molecules (i.e. mono- and disaccharides, pectic sugars, GOS and POS) systematically breaking bonds within the molecule (Allen et al., 2016). All compounds included in this library are listed below:

Mono- and disaccharides from milk and GOS. *Myo*-inositol, galactose, and glucose were included as monosaccharides. For disaccharides, lactose (Gal-β(1,4)-Glc), allolactose (Gal-β(1,6)-Glc), Gal-β(1,3)-Glc, Gal-β(1,3)-Gal, Gal-β(1,4)-Gal and 6-galactobiose (Gal-β(1,6)-Gal) were considered. These GOS structures were previously reported in commercial yogurt and other fermented milks by Toba et al. (1982) and Martínez-Villaluenga et al., (2008).

POS. Possible structures present in homogalacturonan and rhamnogalacturonan-I were taken in to account according to our previous studies (Sabater et al., 2019a; Sabater et al., 2019b).

Then, we assigned a chemical structure for each *m/z* ion by looking for these specific ruptures in our library. This code is able to calculate all possible fragments that can be formed from one molecule (the ones registered in our *in silico* fragmentation library), including both immediate descendants and fragments formed from these descendants. Figure R5.2 illustrates the chemical origin of each fragment, considering that most conformations are possible. Statistically significant differences among these fragments ($p < 0.05$) are discussed below:

Non-specific ions (n=3). Ions ***m/z* 165, 207** and **354** (Figure R5.2a), derived from all GOS and most POS, were more abundant (high) in these two types of oligosaccharides.

High ions in known sugars (n=4). These ions were ***m/z* 322, 324, 336** and **426**. Fragments ***m/z* 322** and **336** (Figure R5.2b) derived from allolactose and 6-galactobiose as well as galactose, glucose and lactose. Similarly, ion ***m/z* 426** is originated from galactose, glucose, allolactose, 6-galactobiose and pectic monosaccharides and ***m/z* 324** (Figure R5.2b) is formed from lactose, allolactose, 6-galactobiose and all identified monosaccharides except *myo*-inositol.

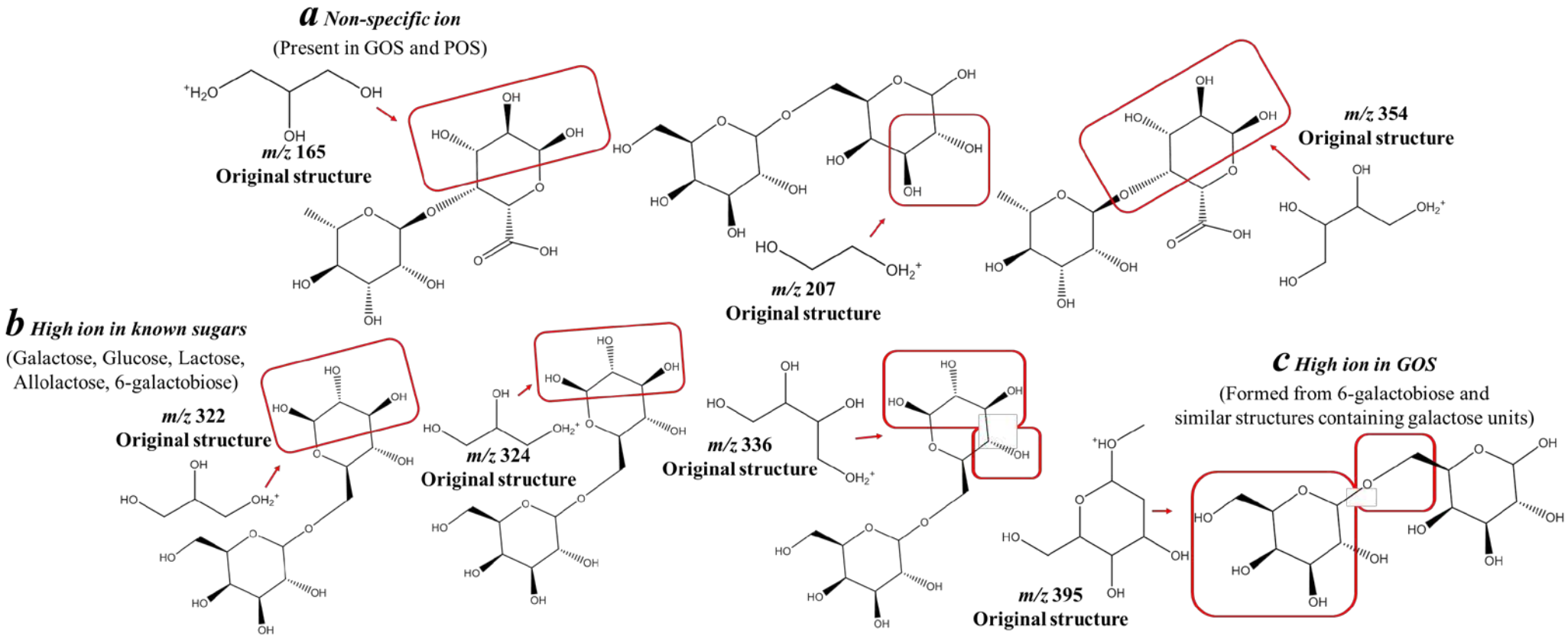


Figure R5.2. Chemical origin of relevant *m/z* ions of carbohydrates found in yogurts supplemented with citrus and artichoke pectins and POS. Non-derivatized original structures (depicted in Supplementary material Table S8) are illustrated. Only one *m/z* ion possible conformation is shown. Some of these structures were more abundant in artichoke POS released during yogurt manufacture (APOS-F).

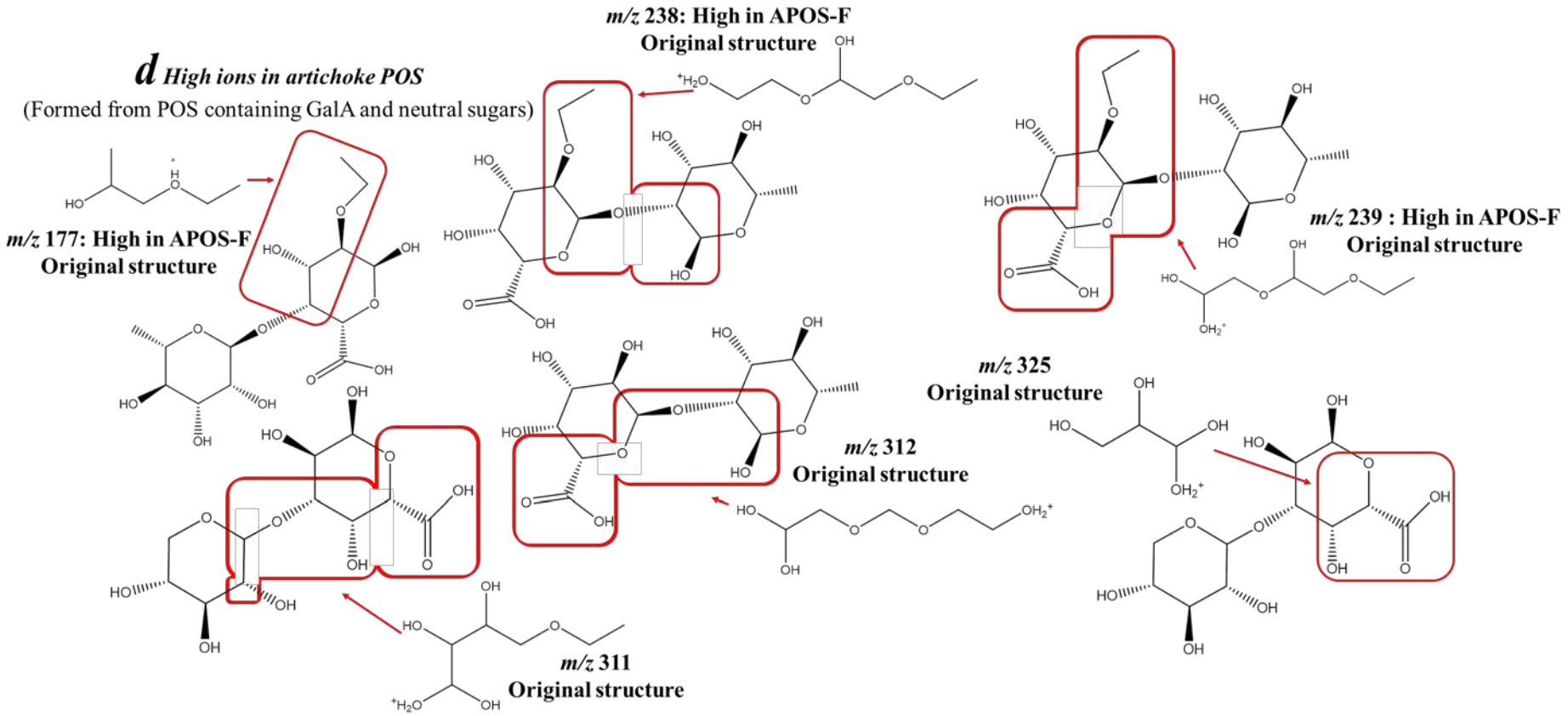


Figure R5.2. Cont.

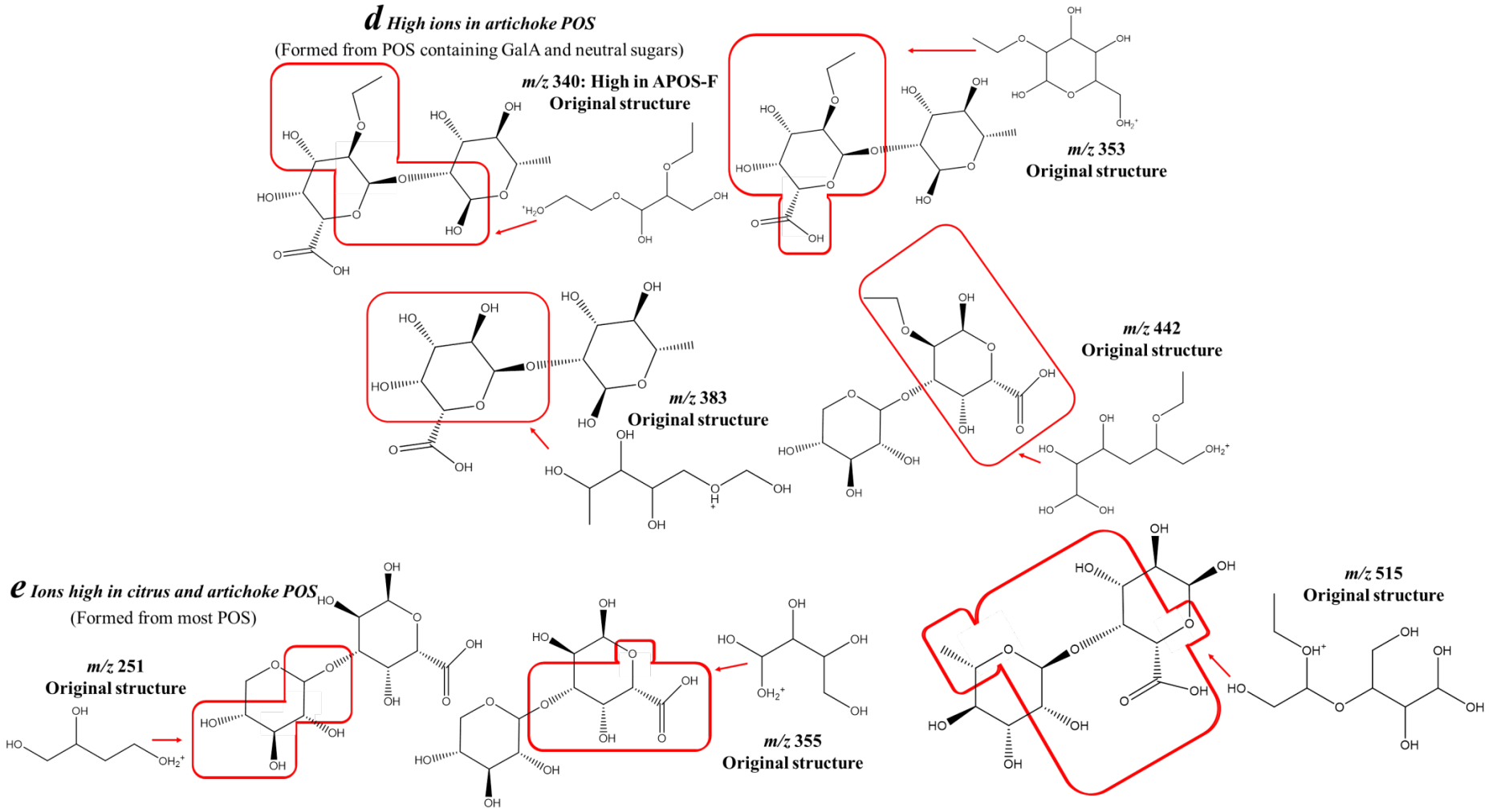


Figure R5.2. Cont.

High ions in GOS (n=1). Ion ***m/z* 395** (Figure R5.2c) formed from 6-galactobiose and similar structures (containing galactose units) was high in GOS.

High ions in artichoke POS (n=14). Fragments **177, 238, 239, 311, 312, 325, 339, 340, 353, 383, 442, 458, 497** and **543** were more abundant in artichoke POS. Chemical origin of some of these fragments (***m/z* 177, 238, 239, 311, 312, 325, 340, 353, 383** and **442**) is shown in Figure R5.2d. Ion ***m/z* 339** is non-specific, ***m/z* 497** is formed from POS ruptures and ***m/z* 311** and **458** are formed from GalA-containing POS. Fragments ***m/z* 325, 353, 383** and ***m/z* 543** derives from several types of POS and some of them could be also formed from other structures such as GOS. On the other hand, ***m/z* 312** (high in APOS-I) and **239** (high in APOS-F) are specifically originated from GalA-α(1,2)-Rha (acetylated or not) and ***m/z* 177** (high in APOS-F) may be formed from acetylated GalA-α(1,2)-Rha and Rha-α(1,4)-GalA (acetylated). In addition, ***m/z* 238** (high in APOS-F) derives from acetylated GalA-α(1,2)-Rha or Xyl-α(1,3)-GalA (acetylated) and ***m/z* 340** (high in APOS-F) is formed from these two POS as well as Rha-α(1,4)-GalA. Similarly, ***m/z* 442** (high in AP and APOS) derives from GalA and acetylated GalA-containing POS. These ions indicate the presence of these POS containing GalA and rhamnose or xylose and many of them showed low abundances in citrus POS obtained with the same enzyme.

High ions in citrus and artichoke POS (n=4). Fragment ***m/z* 251** (Figure R5.2e), originated from all POS and ***m/z* 355** (Figure R5.2e), from most POS except Ara-α(1,4)-Rha; it must be considered that citrus pectin have low arabinose content, and ***m/z* 515** (Figure R5.2e) and **530**, from POS ruptures, were high in both citrus and artichoke POS.

The presence of relevant fragments in artichoke POS derived from GalA and neutral sugar dimers spectra might be due to a higher neutral sugar content in artichoke pectin chains and therefore, higher neutral sugars in its derived POS. Some of these ions were low in POS derived from citrus pectin where GalA contents were significantly higher (87.5%). Among artichoke POS, structures formed by xylose and rhamnose molecules attached to one or more acetylated GalA residues, were released in small amounts during yogurt fermentation and storage. In our previous study (Sabater et al., 2019a) fragmentation patterns of artichoke POS mixtures obtained with Cellulase from *A. niger* suggested the presence of oligomers containing neutral sugars bonded to GalA units. The present work deepens POS characterization indicating that some structures may be formed during yogurt manufacture process.

Quantitative structure-retention time relationships. The next step in our study was to investigate the relationships between GC-EI-MS spectra fragmentation patterns established in *Carbohydrate spectra classification using machine learning algorithms* section and relative retention time (RRT) of carbohydrates present in yogurts. Therefore, we determined those structures which were more abundant in larger oligosaccharide molecules, some of the released during fermentation. RRT values ranged from 0.309 to 1.651 (Supplementary Material Table S5.9). As expected, RRT was low in mono- and disaccharides and higher in those classes including a higher number of tri- and tetrasaccharides (CP-F, CPOS-F, APOS-I and APOS-F). As explained, some of these POS were released during fermentation/storage. BCART and RF were employed again for predicting RRT (RRTp) of each carbohydrate found in yogurt samples considering the 26 fragments selected in the GC-MS spectra classification described in previous section. These models were compared to QRF, a regression-specific algorithm that may yield more consistent and accurate predictions. The quality of each fit was evaluated considering the following parameters: root mean squared error (RMSE), R^2 and mean adjustment error (MAE) (Supplementary Material Table S5.10). Although RF achieved lower RMSE than BCART (0.036 and 0.153, respectively), QRF showed significantly lower RMSE (0.005) and significantly higher R^2 (0.999) than conventional RF when tested on new samples. Considering experimental RRT range and RMSE coefficients, the percentage of error (calculated on RRT) for BCART, RF and QRF was 11.8, 2.8 and 0.4% for the test phase. Differences between RRT and RRTp were calculated (Supplementary Figure S5.5) highlighting QRF predictive power. As explained, RF outperformed BCART (Supplementary Figure S5.6) although these two models yielded similar performance in the classification step (previous section). However, QRF outperformed both models and allows accurately predicting the RRT of milk sugars, GOS and POS based on the abundances of 26 *m/z* fragments. This algorithm showed high reproducibility and accuracy on new samples. To interpret these models, the most influential *m/z* ions in the RRT of all studied carbohydrates have been determined (Supplementary Material Table S5.11). The most influent ions (QRF importance value $\geq$ 35) were selected. Then, QRF was used to generate 3D plots (Figure R5.3) illustrating the effect of *m/z* ions on RRT.

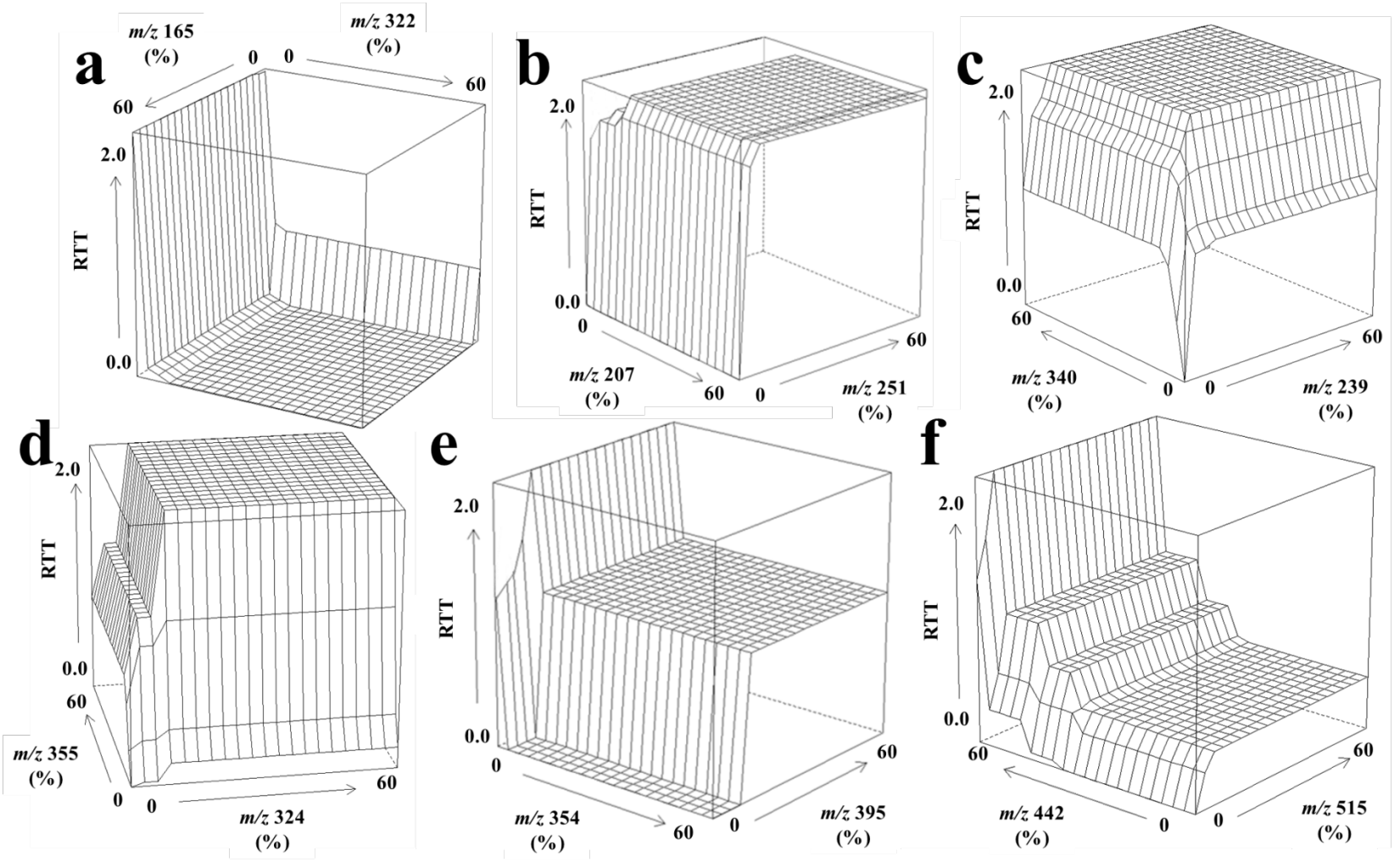


Figure R5.3. 3D plots showing the influence of *m/z* fragment abundances on relative retention time (RRT) of carbohydrates found in all studied yogurts: **a)** *m/z* 165 and 322, **b)** *m/z* 207 and 251, **c)** *m/z* 239 and 340, **d)** *m/z* 324 and 355, **e)** *m/z* 354 and 395, **f)** *m/z* 442 and 515.

Some fragments exerted a positive influence on RRT (***m/z* 207**, **239, 251**, **324**, **340, 355**, **395, 442** and **515**) and high abundances of these ions (up to 60%) contributed to higher RRT values. Among these *m/z* ions, some specific fragments showing higher abundances in larger structures were ***m/z* 442** from acetylated GalA-containing POS as well as ***m/z* 239**, originated from GalA-α(1,2)-Rha (acetylated or not) and ***m/z* 340** from Rha-α(1,4)-GalA, acetylated GalA-α(1,2)-Rha or Xyl-α(1,3)-GalA (acetylated), indicating the presence of this substructures in larger POS molecules with higher RRT (tri- and tetrasaccharides). Some of these ions were high in several artichoke and citrus POS (discussed in previous section). Interestingly, oligosaccharides with the highest RRT were found in POS-supplemented yogurts. It should be noted that ***m/z* 239** and **340** were high in APOS released during fermentation that may correspond to tri- and tetra-POS containing rhamnose or xylose attached to several GalA units (acetylated or not). On the other hand, high abundances of other fragments (***m/z* 165**, **322**, **354**) lead to a decrease in the RRT. These ions are non-specific or derive from milk sugars (mono- and disaccharides), explaining their negative influence of RRT (i. e. more abundant in small molecules). These quantitative structure-retention time relationships give complementary information to the fragmentation patterns described in previous section by determining which specific structures are more abundant in larger oligosaccharide molecules released during fermentation.

Convolutional neural network for establishing structural patterns of functional yogurts carbohydrate fraction. The last step of this GC-EI-MS study focused on stablishing a more confident structural pattern for the complete carbohydrate fraction of these yogurts that confirms the results obtained in the previous steps. Individual GC-EI-MS spectra were classified in first section of this point, to select 26 relevant *m/z* ions whose chemical origin is discussed. Then, the abundance of these structures on larger oligosaccharides, some of them released during yogurt manufacture, was investigated. Finally, a study comparing complete carbohydrate profiles of functional yogurts, and not individual structures present in these samples, was carried out to correlate all information generated. With this aim, heatmaps of the abundances of these 26 relevant *m/z* ions in all carbohydrates present in each type of yogurt have been generated (Figure R5.4). As it can be seen, different profiles were obtained depending on the yogurt. Ion ***m/z* 207**, formed from the two types oligosaccharides studied (GOS and POS), had high abundances in the carbohydrate fraction of all yogurts (up to 60%) at RRT corresponding to di- and trisaccharides. Ion ***m/z* 353**, from most POS and other

oligosaccharides such as GOS, was present in the carbohydrate fraction of all yogurts in abundances between 20-60%. Some di-POS present in POS supplemented with artichoke pectin, showed higher abundances (up to 60%). Similarly, ***m/z* 458** from GalA-containing POS, was significantly abundant (up to 60%) in di-POS from this type of functional yogurt.

On the other hand, yogurts containing artichoke POS showed characteristic structural profiles of di-, tri- and tetra-POS with high abundances of general POS fragments ***m/z* 165, 251** and **311** with abundances up to 20, 10 and 10%, respectively. In addition, several POS showed relevant abundances of specific ions ***m/z* 177** (10-20%) and **239** (20-50%) high in POS released during fermentation as well as ***m/z* 312** (10%) high in initial artichoke POS. These structures may be originated from GalA-α(1,2)-Rha (acetylated or not) and Rha-α(1,4)-GalA (acetylated). Artichoke tri- and tetra-POS also had great abundances of POS-derived ions ***m/z* 339** (5-20%), **353** and **354** (around 20 and 10%, also present in other types of oligosaccharides), **355** and **383** (up to 60%)**.** Moreover, ***m/z* 340** (50-60%) derived from Rha-α(1,4)-GalA**,** acetylated GalA-α(1,2)-Rha or Xyl-α(1,3)-GalA (acetylated) was characteristic of artichoke tetra-POS found in yogurts. Interestingly, ***m/z* 239** and **340** characteristic from POS containing GalA and rhamnose dimers are associated with higher RRT due to their presence in larger oligosaccharides (second section of this point). These fragments were high in APOS-F, released during fermentation. Therefore, artichoke tri- and tetra-POS specific fragmentation patterns associated to POS released during yogurt manufacture have been determined. These structures, released from rhamnogalacturonan-I domain, contain neutral sugars (xylose, rhamnose) attached to several GalA units that may be acetylated.

To mathematically verify these differences in the carbohydrate fraction of functional yogurts, three CNN models were trained on heatmaps to discriminate yogurts containing citrus pectin/POS (CNN1), artichoke pectin/POS (CNN2), APOS from the rest of yogurts (CNN3). Moreover, an additional CNN4 was computed to accurately verify significant structural changes in the complete carbohydrate fraction of functional yogurts during fermentation. Accuracies for all CNN models tested on new samples were 100% at 23, 23, 10 and 38 epochs (full training cycles), respectively. Considering no misclassifications occurred during the train and test phases, model sensitivity, specificity, true positives and negatives rates, and balanced accuracy were 100%.

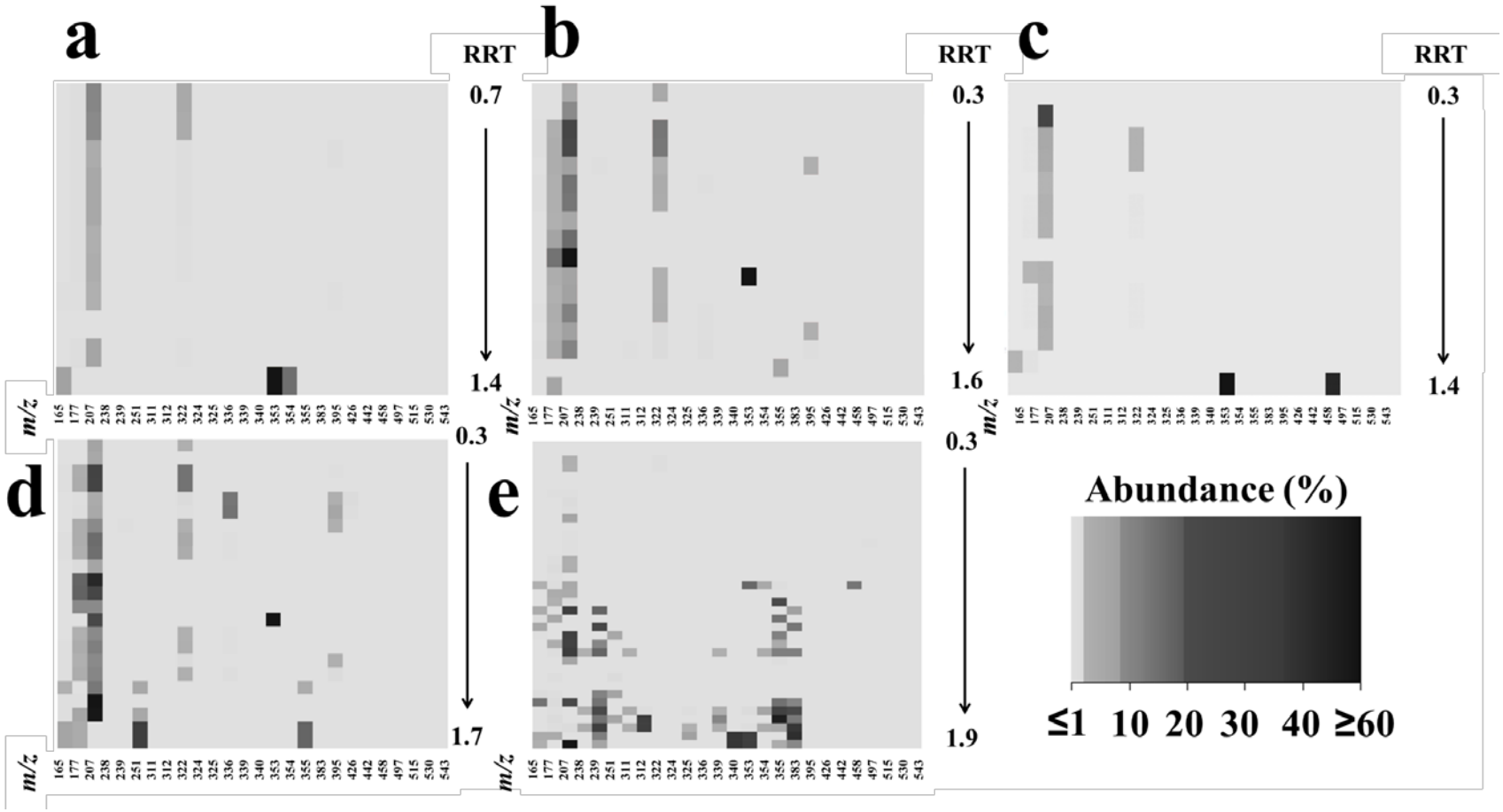


Figure R5.4. Heatmaps showing the abundances of relevant *m/z* ions of carbohydrate fraction from elaborated yogurts used to calculate structure-retention time relationships: **a)** set-style yogurt, and yogurts containing: **b)** citrus and **c)** artichoke pectins and, **d)** citrus and **e)** artichoke POS. RRT: relative retention time.

These models allow summarizing all the information extracted in the previous steps (spectra classification to get characteristic fragments and structure-RRT relationships to illustrate the abundance of these fragments on larger molecules) and highlight differences in the complete carbohydrate fraction of functional yogurts containing functional carbohydrates from two different sources (citrus and artichoke, CNN1 and CNN2). CNN3 also found a specific pattern for APOS determined by the presence of larger oligosaccharide structures released during fermentation formed by several acetylated GalA units attached to one neutral sugar molecule. Furthermore, these structural changes in the carbohydrate fraction of functional yogurts after fermentation were also confirmed by CNN4. This model highlights POS formation during yogurt manufacture and reinforces our results. CNN have been previously used to classify GC-EI-MS data obtaining high prediction rates (Skarysz et al., 2018). In our work, an application of CNN to automatically determine structural patterns of novel oligosaccharide mixtures incorporated in food matrices is presented.

MALDI-TOF-MS analysis to confirm structures elucidated by machine learning models. To confirm the structural features elucidated by machine learning models based on spectral data of a relatively simple technique like GC-EI-MS, yogurt samples were also analysed by MALDI-TOF-MS (Table R5.2). Disaccharides formed by two hexoses are present in all yogurts and correspond to lactose and di-GOS formed. With regard to initial POS present at the beginning of fermentation (0 h) in each yogurt, GalA oligomers (DP 2-3) were found in all functional yogurts. Some of these GalA units were methyl-esterified in yogurts containing citrus POS. Tetrasaccharides containing neutral sugars Xyl-α(1,3)-TriGalA and DiGalA-α(1,2)-Rha-α(1,4)-Gal were also present in all functional yogurts with the exception of the artichoke pectin-supplemented ones. Moreover, artichoke-POS supplemented yogurts showed a wide variety of initial POS structures including disaccharides Xyl-α(1,3)-GalA and GalA-α(1,2)-Rha, in agreement with the complex structural profiles observed for these yogurts in the GC-EI-MS study (Supplementary material Table S5.2).

Table R5.2. Oligosaccharide structures present in functional yogurt determined by MALDI-TOF-MS. These molecules correspond to oligosaccharide Na^+ adducts. **Hex**: hexose, **GalA**: galacturonic acid, **GalA(Met)**: methyl-esterified GalA, **GalA(Acet)**: acetylated GalA, **Xyl**: xylose, **Rha**: rhamnose, **Gal**: galactose, **POS**: pectic oligosaccharides, **CP**: yogurts with citrus pectin, **AP**: yogurts with artichoke pectin, **CPOS**: yogurts with citrus POS, **APOS**: yogurts with artichoke POS.

	Set-style			**CP**			**AP**	
m/z	**0h**	**DP**	***m/z***	**0h**	**DP**	***m/z***	**0h**	**DP**
365.1	2x Hex	**2**	**365.1**	2x Hex	**2**	**365.1**	2x Hex	**2**
			393.3	2x GalA	**2**	**393.3**	2x GalA	**2**
			569.4	3x GalA	**3**	**569.4**	3x GalA	**3**
			701.5	3x GalA + 1x Xyl	**4**			
				2x GalA + 1x Rha + 1x Gal	**4**			
	5h			**5h**			**5h**	
			349.3	1x GalA + 1x Xyl	**2**	**349.3**	1x GalA + 1x Xyl	**2**
			363.3	1x GalA + 1x Rha	**2**	**363.3**	1x GalA + 1x Rha	**2**
365.1	2x Hex	**2**	**365.1**	2x Hex	**2**	**365.1**	2x Hex	**2**
			393.3	2x GalA	**2**	**393.3**	2x GalA	**2**
			525.4	2x GalA + 1x Xyl	**3**			
			569.4	3x GalA	**3**	**569.4**	3x GalA	**3**
			701.5	3x GalA + 1x Xyl	**4**			
				2x GalA + 1x Rha + 1x Gal	**4**			
			787.5	3x GalA + 1x GalA(Acet)	**4**			
	1 week			**1 week**			**1 week**	
			349.3	1x GalA + 1x Xyl	**2**	**349.3**	1x GalA + 1x Xyl	**2**
			363.3	1x GalA + 1x Rha	**2**	**363.3**	1x GalA + 1x Rha	**2**
365.1	2x Hex	**2**	**365.1**	2x Hex	**2**	**365.1**	2x Hex	**2**
			393.3	2x GalA	**2**	**393.3**	2x GalA	**2**
			525.4	2x GalA + 1x Xyl	**3**			
			569.4	3x GalA	**3**	**569.4**	3x GalA	**3**
			701.5	3x GalA + 1x Xyl	**4**			
				2x GalA + 1x Rha + 1x Gal	**4**			
			787.5	3x GalA + 1x GalA(Acet)	**4**			

Table R5.2. Cont.

CPOS			APOS		
m/z	**0h**	**DP**	***m/z***	**0h**	**DP**
			349.3	1x GalA + 1x Xyl	**2**
			363.3	1x GalA + 1x Rha	**2**
365.1	2x Hex	**2**	**365.1**	2x Hex	**2**
393.3	2x GalA	**2**	**393.3**	2x GalA	**2**
569.2	3x GalA	**3**	**569.2**	3x GalA	**3**
583.2	2x GalA + 1x GalA(Met)	**3**			
701.5	3x GalA + 1x Xyl	**4**	**701.5**	3x GalA + 1x Xyl	**4**
	2x GalA + 1x Rha + 1x Gal	**4**		2x GalA + 1x Rha + 1x Gal	**4**
5h			**5h**		
			349.3	1x GalA + 1x Xyl	**2**
			363.3	1x GalA + 1x Rha	**2**
365.1	2x Hex	**2**	**365.1**	2x Hex	**2**
393.3	2x GalA	**2**	**393.3**	2x GalA	**2**
			567.5	1x GalA + 1x GalA(Acet) + 1x Xyl	**3**
569.2	3x GalA	**3**	**569.2**	3x GalA	**3**
			581.5	1x GalA + 1x GalA(Acet) + 1x Rha	**3**
583.2	2x GalA + 1x GalA(Met)	**3**			
611.5	2x GalA + 1x GalA(Acet)	**3**	**611.5**	2x GalA + 1x GalA(Acet)	**3**
701.5	3x GalA + 1x Xyl	**4**	**701.5**	3x GalA + 1x Xyl	**4**
	2x GalA + 1x Rha + 1x Gal	**4**		2x GalA + 1x Rha + 1x Gal	**4**
			743.6	2x GalA + 1x GalA(Acet) + 1x Xyl	**4**
			757.4	2x GalA + 1x GalA(Acet) + 1x Rha	**4**
787.5	3x GalA + 1x GalA(Acet)	**4**	**787.5**	3x GalA + 1x GalA(Acet)	**4**
			877.5	4x GalA + 1x Xyl	**5**
				3x GalA + 1x Rha + 1x Gal	**5**
			1151.7	3x GalA + 2x GalA(Acet) + 1x Rha	**6**
1 week			**1 week**		
			349.3	1x GalA + 1x Xyl	**2**
			363.3	1x GalA + 1x Rha	**2**
365.1	2x Hex	**2**	**365.1**	2x Hex	**2**
393.3	2x GalA	**2**	**393.3**	2x GalA	**2**
			567.5	1x GalA + 1x GalA(Acet) + 1x Xyl	**3**
569.2	3x GalA	**3**	**569.2**	3x GalA	**3**
			581.5	1x GalA + 1x GalA(Acet) + 1x Rha	**3**
583.2	2x GalA + 1x GalA(Met)	**3**			
611.5	2x GalA + 1x GalA(Acet)	**3**	**611.5**	2x GalA + 1x GalA(Acet)	**3**
701.5	3x GalA + 1x Xyl	**4**	**701.5**	3x GalA + 1x Xyl	**4**
	2x GalA + 1x Rha + 1x Gal	**4**		2x GalA + 1x Rha + 1x Gal	**4**
			743.6	2x GalA + 1x GalA(Acet) + 1x Xyl	**4**
			757.4	2x GalA + 1x GalA(Acet) + 1x Rha	**4**
787.5	3x GalA + 1x GalA(Acet)	**4**	**787.5**	3x GalA + 1x GalA(Acet)	**4**
			877.5	4x GalA + 1x Xyl	**5**
				3x GalA + 1x Rha + 1x Gal	**5**
			1151.7	3x GalA + 2x GalA(Acet) + 1x Rha	**6**

With regard to those POS structures released during fermentation, the same disaccharides Xyl-α(1,3)-GalA and GalA-α(1,2)-Rha were released in yogurts supplemented with citrus and artichoke pectin, while trisaccharide Xyl-α(1,3)-DiGalA and tetrasaccharide TriGalA-α(1,4)-GalA(acetylated) were also released in citrus pectin-containing yogurts. Similarly, GalA oligomers containing acetylated units were also released in citrus POS-supplemented yogurts: tri- and tetrasaccharides DiGalA-α(1,4)-GalA(acetylated) and TriGalA-α(1,4)-GalA(acetylated).

Interestingly, in those yogurts containing artichoke POS a wide variety of different structures were released during fermentation. Some of these olgiosaccharides include Tri- and tetra-POS DiGalA-α(1,4)-GalA(acetylated) and TriGalA-α(1,4)-GalA(acetylated) similar to those released in yogurts supplemented with citrus pectin/POS. In addition, Tri- and tetra-POS containing acetylated GalA units and xylose, Xyl-α(1,3)-GalA-α(1,4)-GalA(acetylated) and Xyl-α(1,3)-DiGalA-α(1,4)-GalA(acetylated), or rhamnose, GalA-α(1,4)-GalA(acetylated)-α(1,2)-Rha and GalA-α(1,4)-DiGalA(acetylated)-α(1,2)-Rha, were formed during yogurt incubation. In this type of yogurts, larger POS structures (penta- and hexa-POS) containing neutral sugars, Xyl-α(1,3)-TetraGalA and TriGalA-α(1,4)-DiGalA(acetylated)-α(1,2)-Rha, were also released.

It should be noted that no changes in the structural patterns of samples obtained after fermentation and at 1 week of cold storage were observed. In MALDI-TOF-MS, POS with higher DP than those analysed by GC-EI-MS were determined, due to the higher sensibility of this technique. However, machine learning models were able to elucidate complex structural characteristics from GC-EI-MS spectra of POS.

CONCLUSIONS

Functional yogurts containing POS from artichoke and citrus have been elaborated for the first time. Yogurt pH, cell count, and lactic acid, as well as milk carbohydrate content accomplished the quality standards. With regard to minor oligosaccharides, APOS-supplemented yogurts showed important differences in their carbohydrate fraction compared to the rest of yogurts. These yogurts showed the lowest allolactose and 6-galactobiose formation and a significant decrease in GalA and initial POS content. These differences may indicate that starter culture can release new POS, although in small amounts. Structure-RRT relationships and CNN models highlight the

release of larger POS structures, containing acetylated GalA units and neutral sugars like xylose and rhamnose, in artichoke POS-supplemented yogurts. These structures were confirmed by MALDI-TOF-MS, so this technique and GC-EI-MS in combination with machine learning algorithms could be used as complementary techniques to characterize complex oligosaccharide mixtures incorporated in food matrices.

Article VI

6.3.2. Article IV: Intestinal anti-inflammatory effects of artichoke pectin and modified pectin fractions in dextran sulfate sodium model of mice colitis. Artificial neural network modelling of inflammatory markers

Intestinal anti-inflammatory effects of artichoke pectin and modified pectin fractions in dextran sulfate sodium model of mice colitis. Artificial neural network modelling of inflammatory markers

Carlos Sabater,[a] Jose Alberto Molina-Tijeras,[b] Teresa Vezza,[b] Nieves Corzo,[a] Antonia Montilla[a] and Pilar Utrilla[b]

[a]*Instituto de Investigación en Ciencias de la Alimentación CIAL, (CSIC-UAM) CEI (UAM + CSIC), C/ Nicolás Cabrera, 9, E-28049 Madrid (Spain).*

bDepartamento de Farmacología, Centro de Investigaciones Biomédicas en Red –Enfermedades Hepáticas y Digestivas (CIBER-EHD), Centro de Investigación Biomédica, Universidad de Granada, Granada, Spain

A B S T R A C T

Anti-inflammatory properties of artichoke pectin and modified fractions (arabinose and galactose-free) used at two doses (40 and 80 mg kg^{-1}) in mice with colitis induced by dextran sulfate sodium has been investigated. Pro-inflammatory markers TNF-α and ICAM-I expression decreased in groups of mice treated with original and arabinose-free artichoke pectin while IL-1β and IL-6 liberation was reduced only in mice groups treated with original artichoke pectin. A decrease in iNOS and TLR-4 expression was observed for most treatments. Intestinal barrier genes expression was also determined. MUC-1 and Occludin increased in groups treated with original artichoke pectin while MUC-3 expression also increased in arabinose-free pectin treatment. Galactose elimination led to a loss of pectin bioactivity. Characteristic expression profiles were established for each treatment through artificial neural networks showing high accuracy rates (≥90%). These results highlight the potential amelioration of inflammatory bowel disease on mice model colitis through artichoke pectin administration.

Keywords: Anti-inflammatory effect, Mice model colitis DSS, Artichoke pectin, Modified pectin, Artificial neural network

INTRODUCTION

Pectin is widely used as functional ingredient in the industry due to its technological properties. However, in recent years, its biological properties have gained attention (Gerschenson, 2017). This complex polysaccharide is mainly composed of a linear chain of galacturonic acid (GalA) units bonded by α(1,4) linkages called homogalacturonan (HG), which comprises approximately 65% of pectin and may be partially methoxylated at the *O-6* carboxyl and may be *O*-acetylated at *O-2* or *O-3*. Other major domains of pectin are rhamnogalacturonan-I (RG-I) and RG–II. RG-I comprises 20-35% of pectin structure and is made of sequences of disaccharide [-4)-α-D-Gal*p*A-(1,2)-α-L-Rha*p*-(1-]$_n$. Rhamnose (Rha) may be substituted at *O*-4 with linear or branched oligosaccharides of arabinose (arabinans), galactose (galactans) or both (arabinogalactans) showing degrees of polymerization (DP) up to 47 units depending on the pectin source (Broxterman et al., 2017). RG-II constitutes a minor and more complex domain which comprises 10% of pectin. This domain consists of 12 different types of sugars in over 20 different linkages and is largely conserved across plant species (Atmodjo et al., 2013). Pectin technological and biological properties are determined by its structural characteristics such as monomeric composition, molecular weight (M_w), presence and distribution of side chains, degree of methyl-esterification (DM, % methoxylated GalA) and acetylation, and charge distribution along their backbone (Marić et al., 2018).

Several studies evaluated the *in vivo* bioactivity of pectin including its non-pharmacological administration to ameliorate the symptoms of inflammatory bowel disease (IBD) (Jin et al., 2019; Pacheco et al., 2018). One of the most studied models for investigating the pathogenesis of IBD is the dextran sulfate sodium (DSS)-induced mouse colitis model characterized by bloody faeces, diarrhoea, weight loss and tissue inflammation. Expression of pro-inflammatory cytokines (IL-1β, TNF-α and IL-6) that may impair intestine permeability and mucosal barrier function is also increased, leading to intestine inflammation (Zhu et al., 2017). It has been reported that diets containing citrus pectin and probiotics (*Bifidobacterium longum*) ameliorated the

symptoms of acute and chronic DSS-induced colitis in mice (Silveira et al., 2017). Likewise, diets containing dietary pectin and inulin reduced intestinal inflammation and cancer incidence in mice (Kim et al., 2018).

Efforts have been made to correlate pectin structural characteristic to its anti-inflammatory effect. Polysaccharides from ginseng with high GalA, galactose and arabinose contents supressed pro-inflammatory cytokine expression in mice (Song et al., 2018), and soluble dietary fiber containing pectic substances, with high GalA contents and DM, reduced gastric ulcer lesions in rats by 87% (Abboud et al., 2019). Bioactivity of specific pectin domains or modified pectin fractions has been also studied. For example, galacturonan from starfruit showing high contents of arabinans and arabinogalactans exerted antinociceptive and anti-inflammatory properties in mice (Leivas et al., 2016), and enzymatically modified apple polysaccharides, using pectinases, may prevent against colitis associated colorectal cancer in mice (Sun et al., 2018). Moreover, it has been demonstrated that arabinose and galactose contents may determine biological activity of pectin (de Oliveira et al., 2017; Argüeso et al., 2009). In this sense pectin from artichoke by-products, by its high content in arabinose (127 mg g^{-1} of pectin) and galactose (24 mg g^{-1} of pectin) (Sabater et al., 2018a), can be an interesting natural product to analyse its anti-inflammatory properties.

On the other hand, data modelling allows discovering valuable information on complex chemical and biological events. With this aim, machine learning algorithms like artificial neural networks (ANN) have been recently applied to characterize artichoke pectin chains extracted by different methods (Sabater et al., 2020), and to determine complex structural patterns of pectic oligosaccharides obtained from different pectin sources (Sabater et al., 2019a; Sabater et al., 2019b). In addition, these advanced data analysis tools might be employed to develop predictive models of disease course and response to therapy, as well as characterization of disease heterogeneity and drug development for IBD (Olivera et al., 2019).

Therefore, the aim of this study was to investigate the anti-inflammatory effect of artichoke pectin (AP) and modified artichoke pectin fractions (without arabinose, APwA; or galactose, APwG) in a DSS model of experimental colitis in mice. Then, specific patterns in cytokine and intestinal protein expression were established for each type of sample through ANN modelling.

MATERIALS AND METHODS

Analytical standards and reagents

Analytical reference standards such as D-xylose, L-arabinose, L-rhamnose, D-galactose, galacturonic acid (GalA) and phenyl-β-glucoside were purchased from Sigma Aldrich (Steinheim, Germany). Viscozyme®L (endo-1,3(4)-β-glucanase from *Aspergillus aculeatus*) was a generous gift from Novozymes (Bagsvaerd, Denmark). β-galactosidase from *Bacillus circulans* and *endo*-1,5-α-arabinanase from *Aspergillus niger* were acquired from Biocon (Barcelona, Spain) and Megazyme (Bray, Ireland), respectively. Citrus pectin (CP) was acquired from CEAMSA (O'Porriño, Spain). DSS (36–50 kDa) was purchased from MP Biomedicals (Santa Ana, CA, USA). RNAlater® was obtained from Sigma Aldrich (Steinheim, Germany), and Tri-Reagent® was acquired from Thermo Fisher Scientific (Invitrogen, USA). The oligo (dT) primers (Promega, Southampton, UK) and KAPA SYBRsFAST qPCR Master Mix (Kapa Biosystems, Inc., Wilmington, MA, USA) were used to perform the qPCR analyses. Ultrapure water (18.2 MΩ cm, with levels of 1–5 ng mL^{-1} total organic carbon and <0.001 EU mL^{-1} pyrogen) produced in-house with a laboratory water purification system (Milli-Q Synthesis A10, Millipore, Billerica, MA, USA) was used throughout.

Obtainment and analysis of modified artichoke pectin

Artichoke pectin (AP) was previously extracted in our laboratory using enzymatic preparation Celluclast®1.5L (Sabater et al., 2018a). Galactose-free pectin (APwG) was obtained by enzymatic hydrolysis of 2% (w/v) AP solutions dissolved in 0.05 M acetate buffer (pH 5.0) with 0.7 U mL^{-1} of β-galactosidase from *B. circulans* at 50 ºC for 5h. Similarly, arabinose-free pectin (APwA) was obtained by enzymatic hydrolysis of 2% (w/v) AP solutions dissolved in 0.2 M acetate buffer (pH 4.0) with 0.4 U mL^{-1} of *endo*-1,5-α-arabinanase from *A. niger* at 50 ºC for 24h. Samples were immediately immersed in boiling water to inactivate the enzyme. Then, modified pectins were purified using an Ultracel® ultrafiltration (UF) membrane (M_w cut-off 1 kDa; Millipore, Billerica, MA, USA). Retentates were freeze-dried for subsequent analysis.

Analytical determinations

To determine monomeric composition of purified APwG and APwA, samples (2% w/v in 0.05 M sodium acetate buffer; pH 5.0) were hydrolysed with 90 U mL^{-1} of Viscozyme®L preparation and then sugars released were analysed as TMSO by GC-FID, following the method previously described by Sabater et al. (2018a). On the other hand, complete M_w distribution profiles of artichoke modified pectins were determined by HPSEC-ELSD (Sabater et al., 2019a; Sabater et al., 2018a). The DM was determined by FT-IR (Sabater et al., 2020; Sabater et al., 2018a). All analyses were carried out in duplicate.

Experimental design

Potential anti-inflammatory properties in experimental colitis mice of original AP, APwG and APwA were compared to CP which presents recognized anti-inflammatory activity. This study was carried out in accordance with the Guide for the Care and Use of Laboratory Animals as promulgated by the National Institute of Health and the protocols approved by the Ethic Committee of Animal Experimentation of the University of Granada (Spain) (Ref. No. CEEA 2010-286). Male C57BL/6 mice (7–9 weeks old; 23 ± 2 g) were purchased from Janvier (St Berthevin Cedex, France). Mice were housed in Makrolon cages, maintained under an air-conditioned atmosphere with a 12 h light–dark cycle, and provided with free access to tap water and a standard rodent diet (Panlab A04 diet, Panlab S.A., Barcelona, Spain). Mice were maintained under specific pathogen-free conditions in the facilities of the University of Granada and were randomly assigned to ten groups of six animals each (Table R6.1): 1) healthy control, 2) DSS control, 3) CP 40 mg kg^{-1}, 4) CP 80 mg kg^{-1}, 5) AP 40 mg kg^{-1}, 6) AP 80 mg kg^{-1}, 7) APwA 40 mg kg^{-1}, 8) APwA 80 mg kg^{-1}, 9) APwG 40 mg kg^{-1}, 10) APwG 80 mg kg^{-1}. Induction of colitis was performed 15 days after the start of the experiment by adding 3.0% (w/v) DSS to the drinking water for five days (Table R6.1). Pectin fractions were diluted with water and administered by oral gavage (100 µL per day) corresponding to doses of 40 and 80 mg kg^{-1}. Animal body weight, presence of gross blood in the faeces and the stool consistency were registered to calculate the Disease Activity Index (DAI) (Table S6.1). Once the animals were sacrificed, the colon was aseptically removed and washed and then weighed. Ratio between colon weight and length was also considered as a macroscopic indicator. Then, the expressions of pro-inflammatory cytokines and

barrier intestinal proteins were evaluated. Total RNA from colonic samples was isolated, reverse transcribed and amplified by qPCR using specific primers (Table R6.2). mRNA expression was normalized using the housekeeping gene, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as internal control. The mRNA relative was calculated using the ΔΔCt method.

Data analysis

ANOVA tests and Tukey´s test for $p < 0.05$ were applied to all data generated. In order to establish reproducible patterns in the expression of cytokines and intestinal proteins for each group of mice treated with pectins, artificial neural network (ANN) models were developed. ANNs are powerful pattern-recognition algorithms and the most common type of ANN is the multilayer perceptron. This model is formed by an input layer (i.e. biochemical markers), an output layer (i.e. treatment of mice groups) and several neurons or nodes organized in hidden layers, where each neuron in a layer is connected with each neuron in the next layer through a weighted connection. In addition, a mathematical transformation (activation function) is applied to the input layer to determine whether the information that the neuron is receiving is relevant or not. Six ANNs were computed to determine specific patterns for all studied samples (both doses were included in the same sample type group): 1) Healthy control (ANN-1); 2) DSS control (ANN-2); 3) CP (ANN-3); 4) AP (ANN-4); 5) APwA (ANN-5); 6) APwG (ANN-6). Fig. 1 shows the architecture of ANN used to determine patterns in pectin and modified pectin treatments. The numbers of neurons in the hidden layer were 1, 5, 7, 50, 45 and 50 for ANN-1, -2, -3, -4, -5 and -6, respectively. The activation function for ANN-1, -2 and -3 was logistic while rectifier, tanh and maxout functions were used for ANN-4, -5 and -6, respectively. Variables were scaled and centered before computing the analyses. All the models were trained with 70% of the data, 10-fold cross-validated and then tested with 30% of data from each class (corresponding to new samples). Moreover, a variable importance analysis was carried out. All statistical analyses were computed on R v3.5.0. ANN models were built using RSNNS and H_2O packages (LeDell et al., 2019; Bergmeir & Benitez, 2012).

Table R6.1. Experimental conditions used to study the effect of consumption of artichoke pectin (AP), citrus pectin (CP) and modified artichoke pectin fractions (**APwA**: modified artichoke pectin without arabinose and **APwG**: modified artichoke pectin without galactose) on mice treated with **DSS** (dextran sulfate sodium).

Mice Groups (n=6)	**Animal weight (kg)**	**Daily ingredient dose of treatment (mg kg^{-1} of body weight)**	**DSS (3.0%) supply (days)**	**Treatment supply (days)**
Healthy	0.023 ± 0.002	0	-	-
DSS treated	0.022 ± 0.002	0	5	-
DSS treated + CP 40	0.023 ± 0.002	40	5	15
DSS treated + CP 80	0.023 ± 0.001	80	5	15
DSS treated + AP 40	0.023 ± 0.001	40	5	15
DSS treated + AP 80	0.023 ± 0.002	80	5	15
DSS treated + APwA 40	0.023 ± 0.002	40	5	15
DSS treated + APwA 80	0.023 ± 0.001	80	5	15
DSS treated + APwG 40	0.022 ± 0.002	40	5	15
DSS treated + APwG 80	0.023 ± 0.001	80	5	15

Table R6.2. Primer sequences used in real-time qPCR assays in colonic tissue of different groups of mice studied.

Gene	Sequence 5´-3´	Annealing *T* (°C)
GAPDH	FW 5'- CCATCACCATCTTCCAGGAG -3' RV 5'- CCTGCTTCACCACCTTCTTG -3'	60
TNF-α	FW 5'- AACTAGTGGTGCCAGCCGAT -3' RV 5'- CTTCACAGAGCAATGACTCC -3'	60
IL-1β	FW 5'- TGATGAGAATGACCTGTTCT -3' RV 5'- CTTCTTCAAAGATGAAGGAAA -3'	60
IL-6	FW 5'- TAGTCCTTCCTACCCCAATTTCC -3' RV 5'- TTGGTCCTTAGCCACTCCTTCC -3'	60
iNOS	FW 5'- GTTGAAGACTGAGACTCTGG -3' RV 5'- ACTAGGCTACTCCGTGGA -3'	67
ICAM-1	FW 5'- CAGTCCGCTGTGCTTTGAGA -3' RV 5'- CGGAAACGAATACACGGTGAT -3'	60
TLR-4	FW 5'- GCCTTTCAGGGAATTAAGCTCC -3' RV 5'- AGATCAACCGATGGACGTGTAA -3'	60
MUC-1	FW 5'- GCAGTCCTCAGTGGCACCTC -3' RV 5'- CACCGTGGGCTACTGGAGAG -3'	60
MUC-2	FW 5'- GCTGACGAGTGGTTGGTGAAT -3' RV 5'- GATGAGGTGGCAGACAGGAGA -3'	60
MUC-3	FW 5'- CGTGGTCAACTGCGAGAATGG -3' RV 5'- CGGCTCTATCTCTACGCTCTCC -3'	60
Occludin	FW 5'- ACGGACCCTGACCACTATGA -3' RV 5'- TCAGCAGCAGCCATGTACTC -3'	60
ZO-1	FW 5'- GGGGCCTACACTGATCAAGA -3' RV 5'- TGGAGATGAGGCTTCTGCTT -3'	56
TFF-3	FW 5'- CCTGGTTGCTGGGTCCTCTG -3' RV 5'- GCCACGGTTGTTACACTGCTC -3'	60
Villin	FW 5'- CTCCGAGCAGATTGAGAAGG -3' RV 5'- GGTGCTGCCACTCTTCTACC -3'	60

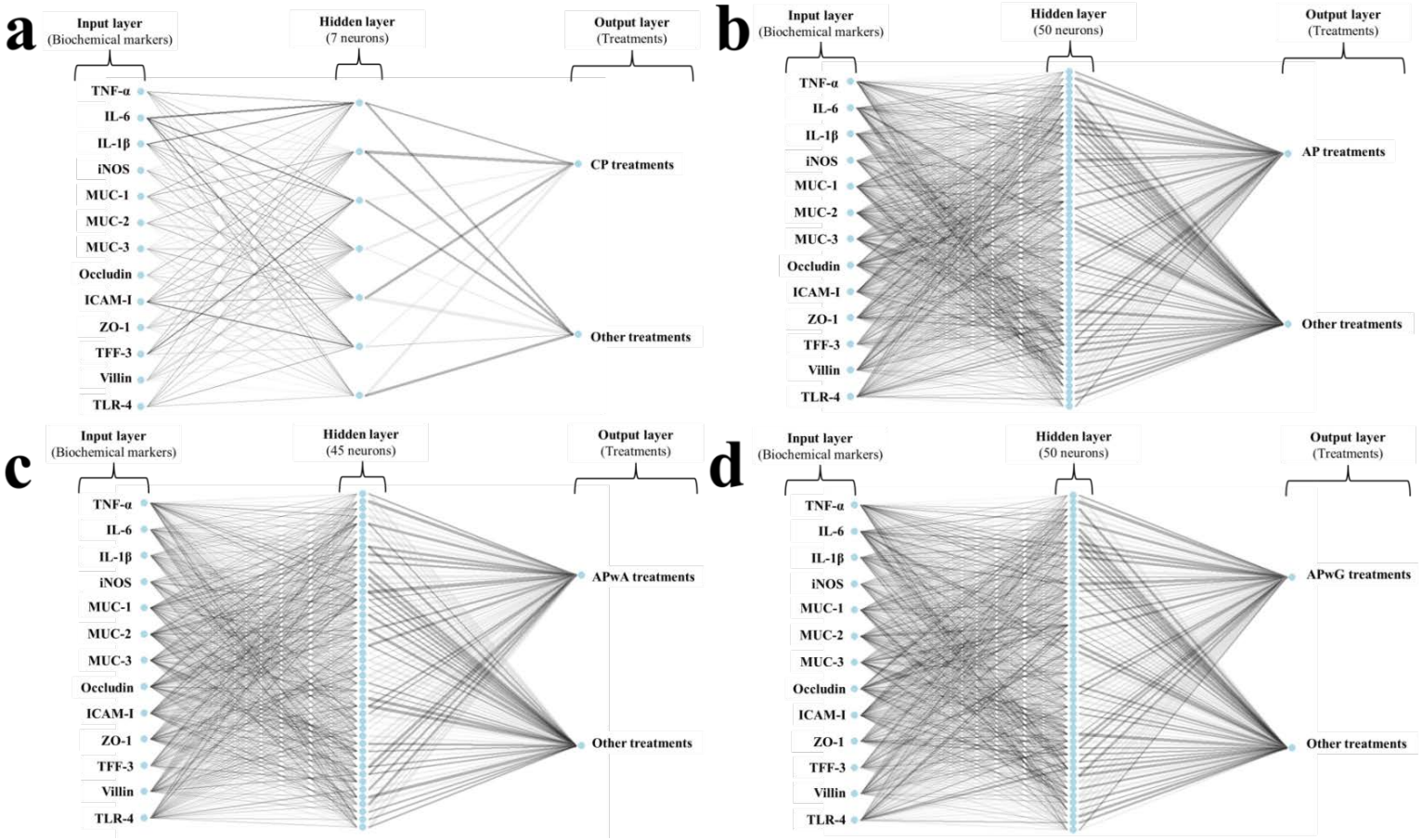


Figure R6.1. Artificial neural network (ANN) architecture used for determining biochemical patterns in mice groups treated with citrus pectin (CP; **a**) and original artichoke pectin (AP); **b**), arabinose-free (APwA; **c**) and galactose-free (APwG; **d**) artichoke pectin. Weights are color-coded by sign (black +, grey -) and thickness is in proportion to magnitude.

RESULTS

Characterization of artichoke pectin and modified pectin fractions

A characterization of CP, AP as well as modified artichoke pectin fractions has been performed in order to know structural differences and thus to assess their influence in each group of mice studied. Enzymatic treatment with β-galactosidase from *B. circulans* and *endo*-1,5-α-arabinanase from *A. niger* released 86.6 and 94.8% of total galactose and arabinose content present in AP structure, respectively. Modified pectin fractions purified by UF showed a higher linearity pectin backbone than original AP, although these differences were not statistically significant. In contrast, APwA showed a significantly lower extent of branching than APwG, indicating the removal of side chains during enzymatic modification (Table S6.2). Arabinose and galactose contents decreased to 1.4 and 0.6% when treated with *endo*-1,5-α-arabinanase from and β-galactosidase from, respectively. Interestingly, these fractions had a slightly higher GalA content (76-77%) than the original pectin chain and APwA showed significantly higher amounts of rhamnose (6.3%) (Table S6.2), highlighting that the linear backbone of HG and RG-I were preserved. Xylose content was significantly higher in APwG (1.5%). In contrast, commercial CP used as control showed significantly higher xylose and galactose contents (2.5 and 11.1%) and lower arabinose contents than AP. Also, it should be considered that CP shows higher DM (71%) than AP (19.5%) and modified pectins.

On the other hand, APwG showed a multimodal M_w distribution similar to original AP (Table S6.3) with three main fragments of 605, 74 and 5.1 kDa. In contrast, APwA had a bimodal distribution showing two main fragments of 542 and 63 kDa, indicating little difference in M_w profiles. It should be noted that CP used as control showed a monomodal distribution with one main fragment of 547 kDa.

Influence of pectin treatment on macroscopic indicators of inflammatory bowel disease in DSS -treated mice

First, different macroscopic indicators and symptoms of colitis induced in mice by the oral administration of DSS were studied. A significant weight loss was observed in all experimental groups except for the healthy control group (Figure R6.2a,b). As expected, DSS-induced colitis produced a weight loss in mice due to a systemic status of illness occasioned by DSS as a general toxic. In general, weight losses in mice

treated with low doses (40 mg kg^{-1}) of artichoke pectin (original and modified) were similar to those of DSS control group. It should be noted that weight losses were more accentuated for mice treated with modified artichoke pectin than original one, indicating a loss of bioactivity. Among modified pectins, a higher weight loss was observed for APwA treated mice groups than those treated with APwG, indicating its lower beneficial effect. Interestingly, the administration of original AP resulted in lower weight reduction than CP with recognized anti-inflammatory activity used as control. This positive effect was especially relevant at doses of 80 mg kg^{-1}, in this group no weight losses were observed at the seventh day of treatment. Disease progression was monitored by calculating DAI for each group (Figure R6.2c,d). DAI values were high for AP and modified pectin administered at doses of 40 mg kg^{-1}, and was even higher for APwA in agreement with high weight loss values. Similarly, AP at doses of 80 mg kg^{-1} resulted in a significant decrease of DAI value on the seventh and eighth days compared to the sixth; these values were lower than the one of CP at the same doses. This fact may indicate a dose-dependent effect. On the other hand, the weight/length ratio of the colon was determined (Figure R6.3). This ratio reached its highest and lowest values in DSS and healthy control groups, respectively. A statistically significant decrease in this ratio ($p < 0.05$) with respect to DSS control group was only obtained for CP, AP and APwG at doses of 80 mg kg^{-1}. These results may highlight a lower severity of inflammation and minor colonic cell infiltration in these groups (CP 80, AP 80 and APwG 80). These results were in agreement with those of Pacheco et al. (2018) who observed a significant reduction of IBD symptoms and improvement of the animals' general status in mice treated with this CP. However, differences between the same product administered at different doses (40 and 80 mg kg^{-1}) were not statistically significant. In contrast, no significant differences were found between DSS group, the groups intake 40 mg kg^{-1} dose of all ingredients and also APwA at 80 mg kg^{-1}.

Considering this set of macroscopic indicators, an anti-inflammatory effect for original AP is shown leading to a significant reduction of symptoms. However, enzymatic modification of AP did not enhance its bioactivity.

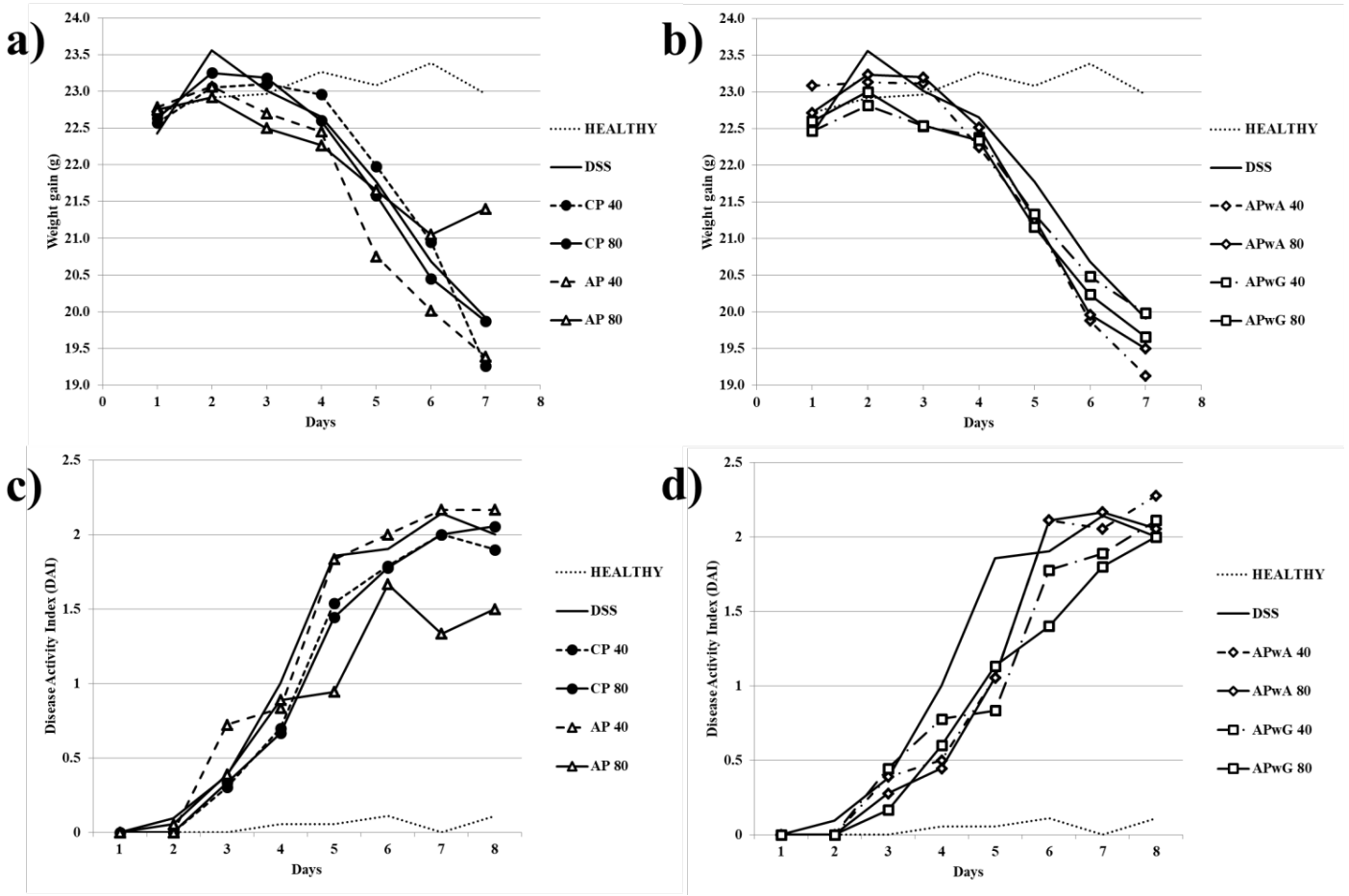


Figure R6.2. Macroscopic indicators weight gain (**a, b**) and Disease Activity Index (DAI) (**c, d**) determined after sacrifice of mice groups: control (healthy), DSS treated (dextran sulfate sodium) and DSS treated + different types of pectins, citrus pectin, (CP), and artichoke pectin, (AP), modified artichoke pectin without arabinose (APwA), and modified artichoke pectin without galactose (APwG) Mice were treated with two doses 40 and 80 mg kg^{-1}.

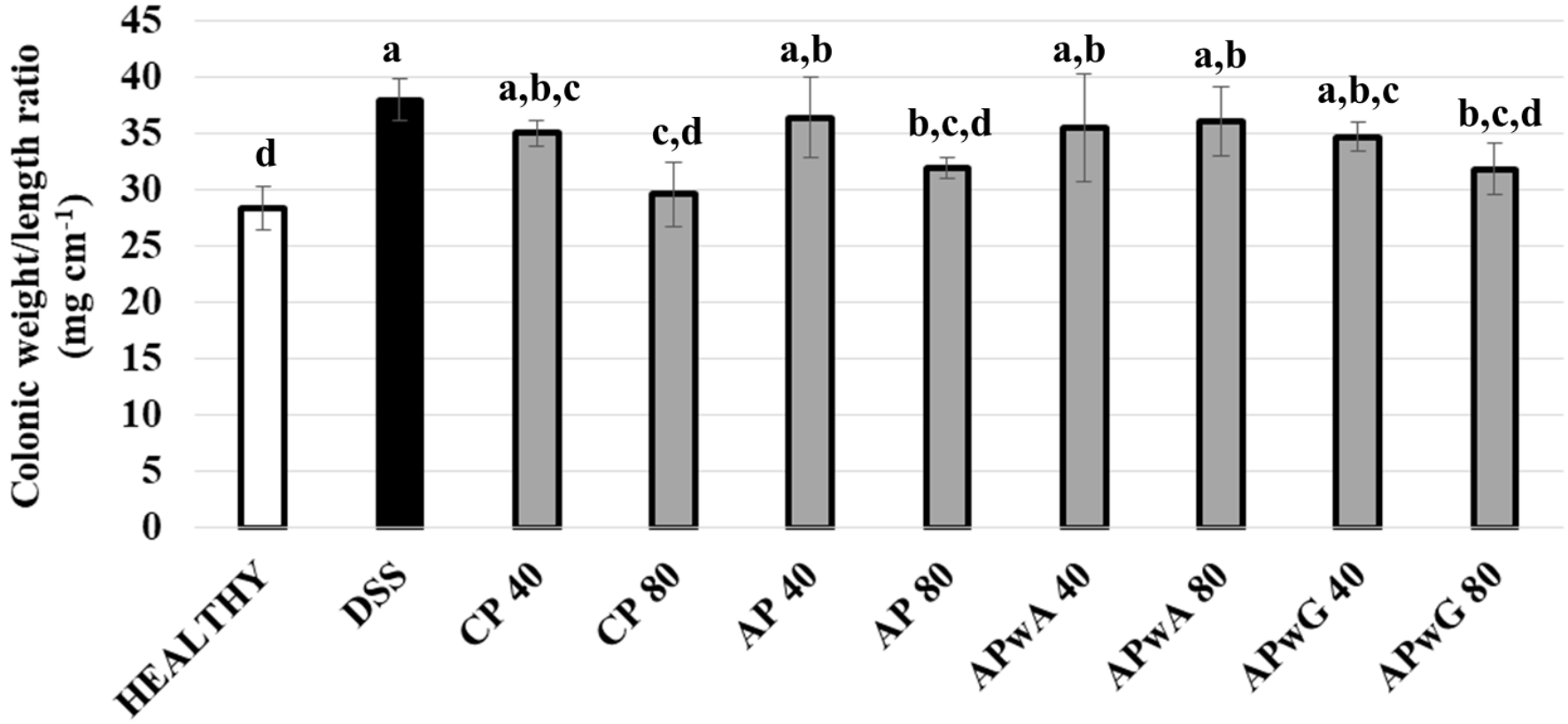


Figure R6.3. Weight/length ratio of the colon determined after sacrifice of mice groups: control (healthy), DSS treated (dextran sulfate sodium) and DSS treated + different types of pectins, citrus pectin, (CP), and artichoke pectin, (AP), modified artichoke pectin without arabinose (APwA), and modified artichoke pectin without galactose (APwG) Mice were treated with two doses 40 and 80 mg kg^{-1}. ^{a,b,c,d} Statistically significant differences between mice groups.

Influence of pectin treatment on inflammatory biochemical markers in DSS -treated mice corroborated by ANN modelling.

To gain deeper knowledge of the disease progression in mice, several biochemical markers of IBD were assessed (Figures R6.4 and R6.5). This set of markers consisted in several pro-inflammatory parameters such as cytokines (TNF-α, IL-1β, IL-6) (Figure R6.4), enzymes (iNOS) and transmembrane receptors (ICAM-I, TLR-4; Figure R6.4); as well as different epithelial integrity markers such as barrier intestinal proteins (MUC-1, MUC-2, MUC-3, Occludin, ZO-1, TFF-3, Villin) (Figure R6.5). Given the great variability observed between individuals, ANN models were trained on biochemical parameters to obtain more robust and confident expression patterns for each group of mice studied (healthy, DSS treated, and DSS treated and supplemented with CP, AP, APwA and APwG that may help to interpret these results. These algorithms found a reproducible classification pattern for each kind of substrate (Table S6.4), showing high prediction rates on the independent test set, above 90% in all cases. Kappa, a more robust accuracy metric that considers the possibility of a correct classification by chance, showed the lowest values for APwA and original AP probably due to the highest differences observed between the effects of these samples at the two studied doses (Figures R6.4 and R6.5). However, kappa values were above 0.80, indicating a high reproducibility. Therefore, ANN was able to find general patterns that allow discriminating the biochemical profile obtained for each group of mice studied considering these high prediction rates, sensitivity, specificity and balanced accuracy of these models (above 90%) (Table S6.4). To interpret ANN results, the importance of each biochemical parameter in these models were calculated (Table R6.3). Differences observed in cytokine expression profiles, based on this analysis, are discussed below.

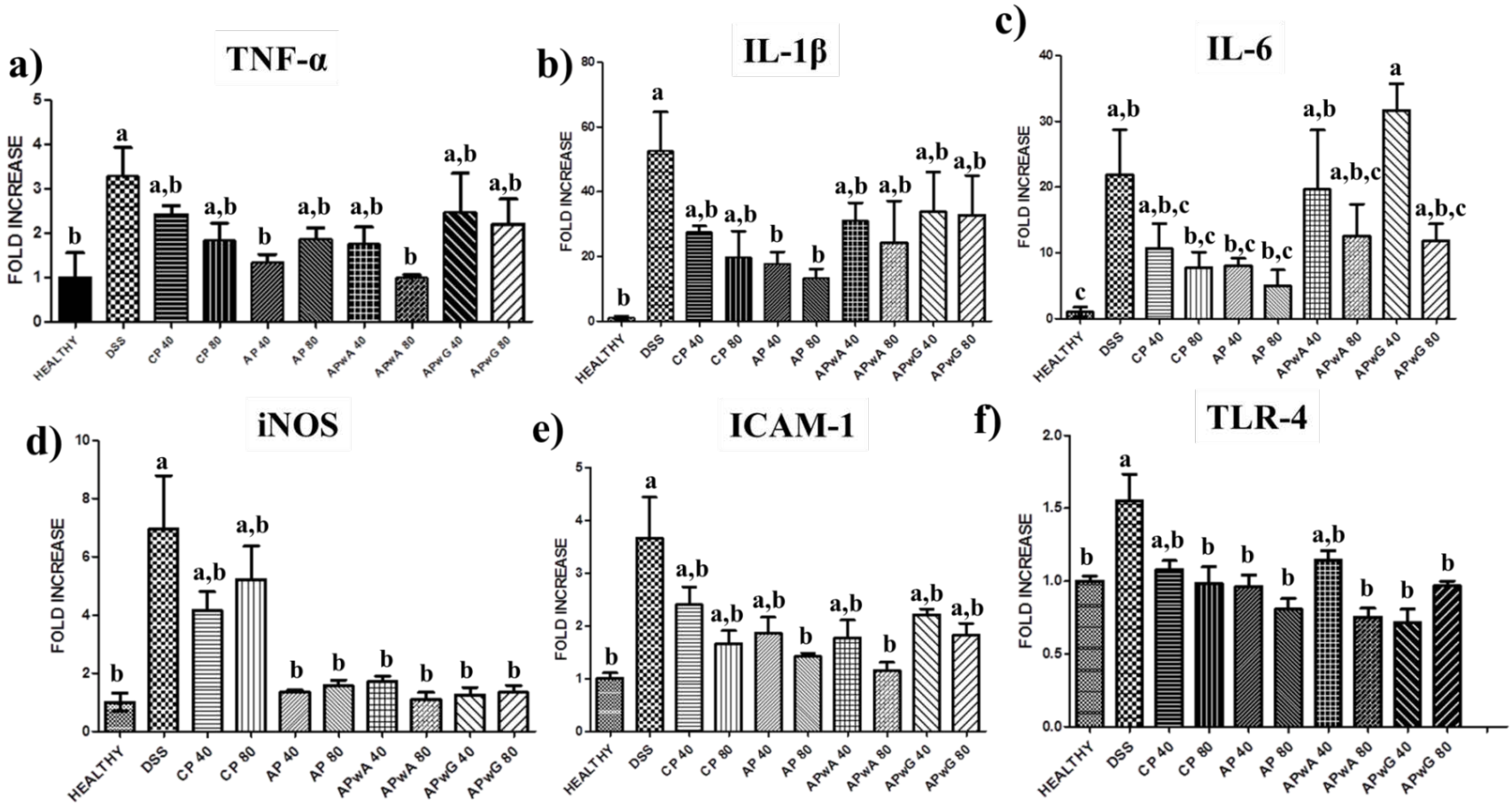


Figure R6.4. Biochemical pro-inflammatory markers determined after sacrifice of mice groups: control (healthy), DSS treated (dextran sulfate sodium) and DSS treated + different types of pectins, citrus pectin, (CP), and artichoke pectin, (AP), modified artichoke pectin without arabinose (APwA), and modified artichoke pectin without galactose (APwG) Mice were treated with two doses 40 and 80 mg kg^{-1}. [a,b,c] Statistically significant differences between mice groups.

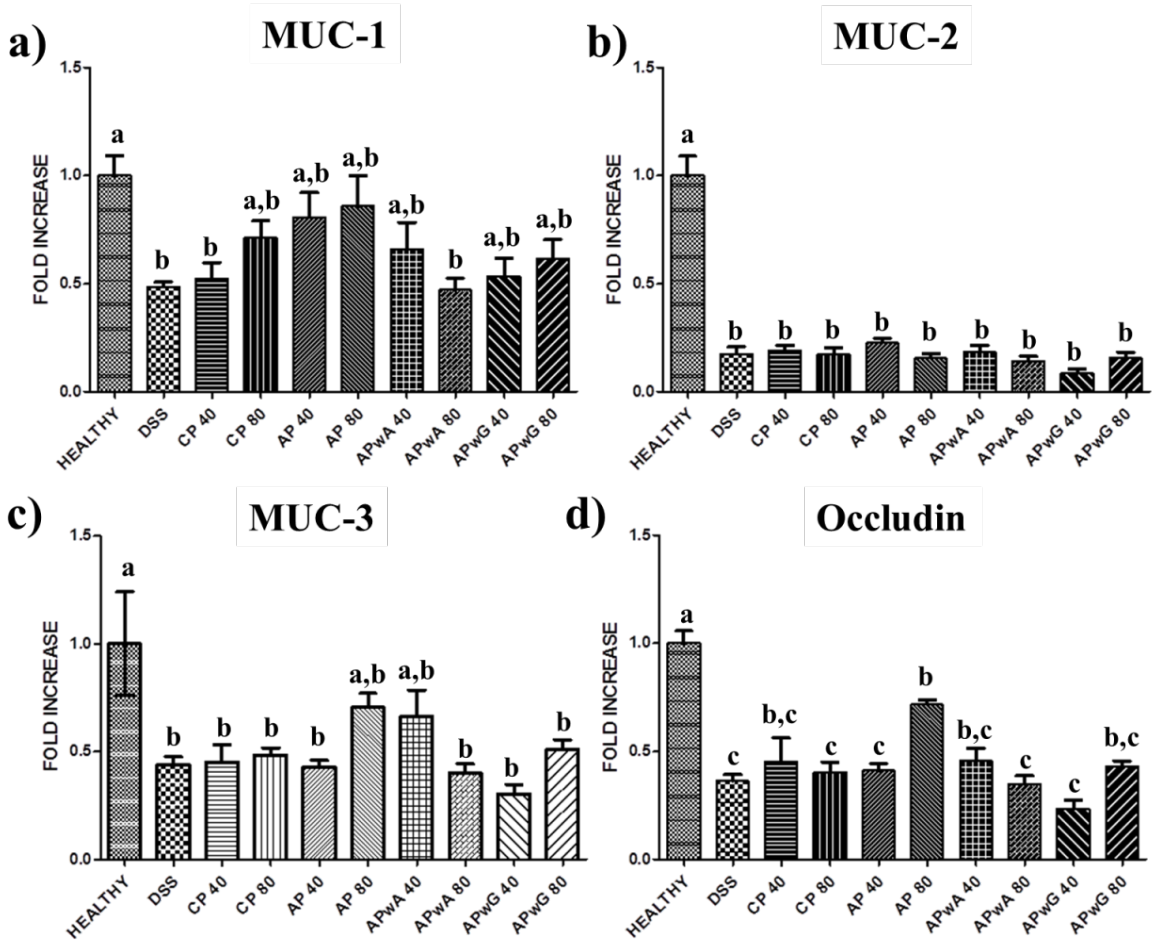


Figure R6.5. Biochemical barrier intestinal proteins determined after sacrifice of mice groups: control (healthy), DSS treated (dextran sulfate sodium) and DSS treated + different types of pectins, citrus pectin, (CP), and artichoke pectin, (AP), modified artichoke pectin without arabinose (APwA), and modified artichoke pectin without galactose (APwG) Mice were treated with two doses 40 and 80 mg kg^{-1}. ^{a,b,c} Statistically significant differences between mice groups.

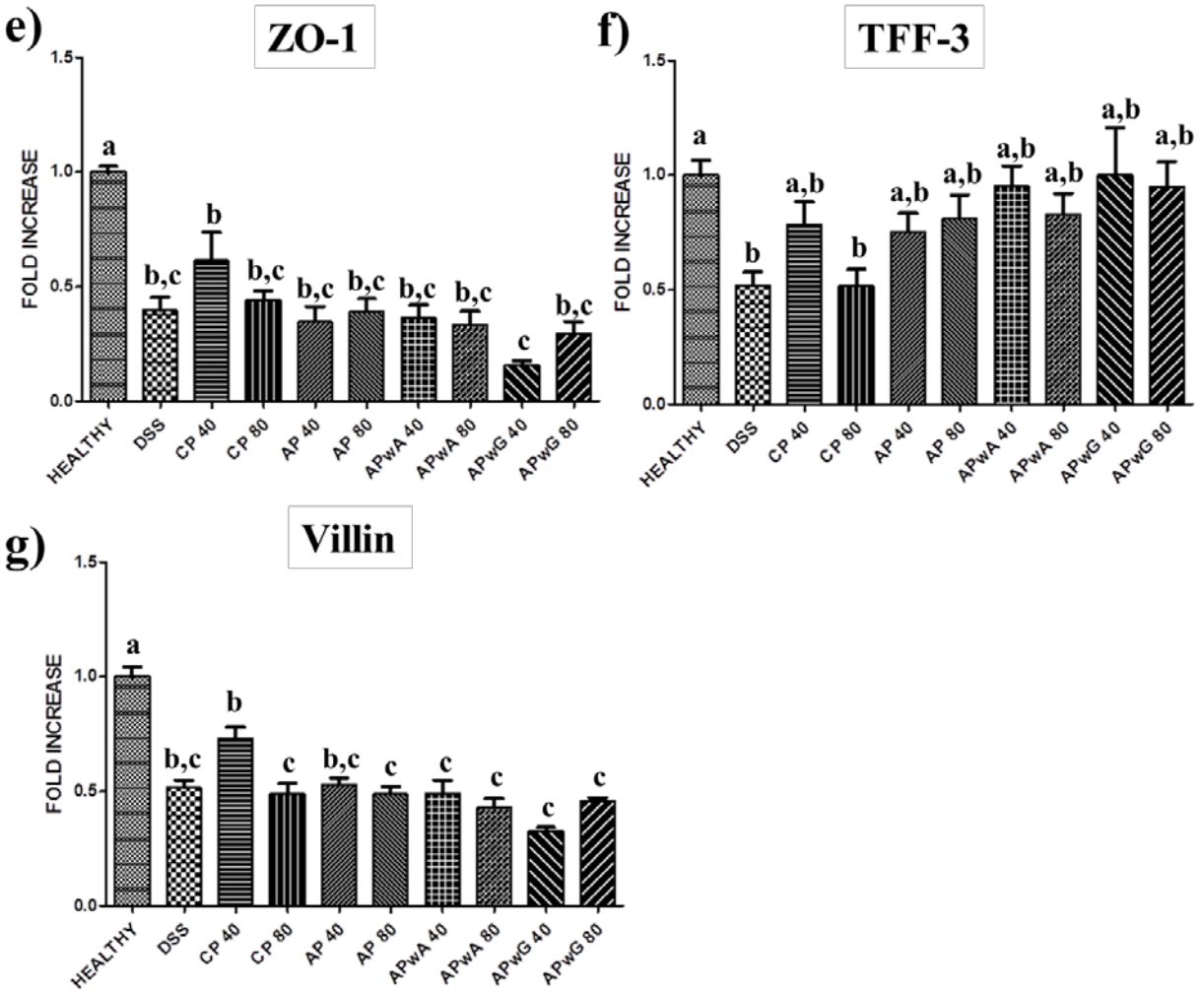


Figure R6.5. Cont.

Table R6.3. Variable importance analysis for all artificial neural network (ANN) model to study characteristic expression profiles of cytokine and intestinal proteins in the different groups of mice studied. **DSS** treated **(**dextran sulfate sodium), **CP**: citrus pectin control, **AP**: artichoke pectin, **APwA**: modified artichoke pectin without arabinose, **APwG**: modified artichoke pectin without galactose.

Parameter	ANN-1 (Healthy control)	DSS treatment				
		ANN-2 (Control)	ANN-3 (+ CP)	ANN-4 (+ AP)	ANN-5 (+ APwA)	ANN-6 (+ APwG)
TNF-α	3.8	**34.0**	46.3	69.3	**96.1**	81.8
IL-6	12.5	15.7	**100**	**100**	92.4	**100**
IL-1β	13.6	15.0	32.3	63.8	59.4	61.4
iNOS	5.4	3.6	**80.8**	63.1	82.9	77.4
ICAM-I	8.2	**100**	9.6	71.5	**100**	**92.1**
TLR-4	23.9	**53.7**	24.7	75.9	87.3	86.4
MUC-1	1.5	21.4	14.2	**80.2**	81.5	86.1
MUC-2	**100**	23.8	26.0	72.1	89.1	86.8
MUC-3	8.2	15.5	27.2	79.8	77.0	76.1
Occludin	**43.9**	7.7	32.0	**86.2**	82.9	**91.0**
ZO-1	33.1	4.9	25.0	75.6	91.8	**87.9**
TFF-3	41.0	27.6	25.6	74.1	76.1	76.6
Villin	**63.2**	22.4	**64.6**	74.3	71.5	77.0

As expected, DSS-induced colitic inflammation was characterized by an altered immune response, which was evidenced by an increase mRNA expression of the pro-inflammatory cytokines in comparison with non-colitic mice (Figure R6.4). TNF-α expression decreased in colitis mice treated with original or modified pectin although this reduction was only statistically significantly ($p < 0.05$) in those groups treated with original AP and APwA pectin at doses of 40 and 80 mg kg^{-1}, respectively, showing most similar values to those of the healthy group. No significant differences were observed between APwG fractions at both doses. ANN analysis corroborated these results; TNF-α was one of the most relevant parameter in the profiles of mice treated with APwA, showing a higher reduction compared to other mice groups (Table R6.3). In addition, TNF-α was especially relevant to discriminate the profiles of DSS control by ANN, indicating that this marker showed an accentuated decrease in experimental mice groups treated with most pectin samples. Similarly, IL-1β was significantly reduced when AP was administered at 40 and 80 mg kg^{-1} while no significant differences were observed among the rest of groups. IL-6 levels in CP and AP treatments at the two studied doses (40 and 80 mg kg^{-1}), were reduced, although differences between groups were not statistically relevant due to the biological variability. It should be noted that APwA and APwG did not reduce the expression of IL-6 at 40 mg kg^{-1}, showing similar values to those of DSS group, indicating a loss of potential immunostimulatory activity. IL-6 was also the most relevant parameter in ANN profiles for CP control, original AP and APwG (Table R6.3). ANN corroborates that CP and original AP treatments reduced this pro-inflammatory cytokine. In the case of APwG, analysis of variable importance indicates the negative results obtained for this type of pectin did not exert any relevant effect on IL-6 levels.

Remarkably, iNOS expression was significantly decreased in all groups treated with AP and modified artichoke pectin fractions at both doses, reaching expression levels similar to those of the healthy group. ANN analysis revealed that iNOS was an important parameter to discriminate the profiles of mice treated with CP (Table R6.3) because this was the only treatment that did not resulted in a reduction of iNOS levels. This fact may be attributed to structural differences found between the two pectins studied, thus CP presented a higher DM (71%) than AP (19.5%) and modified artichoke pectins (24%) (Table S6.2).

Other pro-inflammatory marker determined was transmembrane receptor involved in the inflammatory process ICAM-I (Figure R6.4e), which was significantly reduced

when AP and APwA were administered at doses of 80 mg kg^{-1}, reaching values similar to those of the healthy control. ANN analysis showed that ICAM-I was the most relevant marker to discriminate the biochemical profiles of DSS control groups because all pectin and modified pectin treatments led to a decrease in the levels of this parameter (Table R6.3). ICAM-I was also the most important parameter in the profiles of mice treated with APwA pectin as it was drastically reduced. Interestingly, ICAM-I was relevant in APwG groups because this treatment led to a lower reduction in the levels of ICAM-I compared to other pectin samples.

A decrease in TLR-4 (Figure R6.4f) expression was observed when pectin or modified pectin was administered to mice groups being significant at all conditions except for CP and APwA treatments at 40 mg kg^{-1}. ANN indicated that TLR-4 levels differentiated DSS control from the rest of mice groups (Table R6.3) and corroborated the overall in the expression of TLR-4 produced by all pectin and modified samples,

In general, a reduction of all pro-inflammatory markers is produced, although the differences are not significant in several cases. ANN models confirmed the potential activity of AP and APwA fractions to modulate cytokine expression.

On the other hand, a group of anti-inflammatory markers such as the expression of barrier intestinal proteins were studied (Figure R6.5). MUC-1expression increased in those groups treated with original AP. However, a great variability between individuals was observed. ANN analysis showed that MUC-1 was an important factor in the profiles of original AP treated groups (Table R6.3) as it produced a relevant increase in the levels of this protein, showing the highest expression values.

In contrast, MUC-2 (Figure R6.5b) levels were drastically low in all studied groups with respect to the healthy control, showing values similar to those of DSS control groups, so no positive effect on this marker was observed. Therefore, this was the most relevant marker to discriminate the profiles of healthy mice group by ANN (Table R6.3). Similarly, the expression of MUC-3 (Figure R6.5c) for all tested groups was significantly lower than in healthy group, with the exception of original AP and APwA treatments at 80 and 40 mg kg^{-1}, respectively.

With regard to other membrane proteins determined, Occludin (Figure R6.5d) levels were only significantly high in the group treated with AP at 80 mg kg^{-1}. ANN patterns indicated that Occludin was a relevant parameter in healthy mice group and those mice groups treated with original AP (Table R6.3), as no relevant increases were observed in the levels of Occludin when other pectin samples were administered. Moreover, this

parameter was also relevant to discriminate mice groups treated with APwG highlighting the absence of any positive effect; these groups showed the lowest expression of this barrier protein compared to other pectin treatments.

In addition, no relevant differences were observed in TFF-3 levels (Figure R6.5f) between different pectin treatments due to the high variability between individuals. The expression of ZO-1 and Villin (Figures R6.5e and g) were increased only for CP treatments at 40 mg kg^{-1} and no significant differences were observed between modified pectin fractions and original AP. These results were corroborated by ANN analysis, revealing that Villin was an important parameter in healthy mice group and mice treated with CP (Table R6.3). This fact indicates that treatment of mice groups with CP was one of the most effective for reducing inflammatory cytokine IL-6 (as explained above) leading to an increase the expression of barrier protein Villin. In contrast, ZO-1 was a relevant marker in the biochemical profiles of mice treated with APwG.

As explained, some of these differences were not statistically significant due to the high interindividual variability. However, ANN modelling corroborates differences observed in obtained results and allows determining the most relevant biochemical markers to be considered in each mice group. Therefore, it was observed a response dose-dependent for several biochemical markers: pro-inflammatory TNF-α and IL-1β, intestinal proteins and receptors MUC-1, Occludin and TLR-4. AP was especially efficient to reduce the expression of pro-inflammatory cytokine IL-6 compared to other samples, while APwA produced the most relevant effect on pro-inflammatory receptor ICAM-I decrease. However, galactose-free modification resulted in a loss of pro-inflammatory cytokine regulation. In general, for APwG, the variable importance analysis highlights the negative results obtained for this type of modified pectin, so ANN model determined higher levels of inflammatory markers ICAM-I and IL-6 and lower expression of intestinal barrier proteins Occludin and ZO-1 in the profiles of these groups compared to the rest of tratments. Therefore, enzymatic modification with β-galactosidase from *B. circulans* produced a significant loss of pectin bioactivity.

DISCUSSION

Acute inflammatory response involves high expression of macrophage-derived cytokine profiles with a high participation of TNF-α, IL-1β and iNOS as well as adhesion molecule ICAM-I that enhances leukocyte endothelial transmigration, leading

to tissue damage (Pacheco et al., 2018). Several studies reported *in vitro* and *in vivo* bioactivity of pectin and pectic substances highlighting the importance of specific structural features. Zhang et al. (2019) reported that pectic substances with high arabinose contents inhibit the production of IL-6 and TNF-α in LPS-stimulated RAW264.7 cells while a pectic polysaccharide from alfalfa, consisting mainly in RG-I domain with L-arabinosyl and D-galactosyl units, showed potential inhibition of IL-1β, IL-6 gene expressions in the same cell line (Chen et al., 2015). Arabinogalactan from edible jambo fruit mainly composed of galactose and arabinose attenuated the pro-inflammatory secretion induced by an inflammatory agent in THP-1 cells (Tamiello et al., 2018), and silver linden flower pectins with RG-I domains rich in arabinogalactans supressed iNOS expression and showed macrophage-stimulating activities (Georgiev et al., 2017a), indicating that neutral sugar content of pectin determines its anti-inflammatory activity. These studies highlight the importance of arabinose and galactose contents in pectin potential bioactivity. However, in our study galactose content in AP proved to be especially relevant to preserve pectin bioactivity while APwA conserved its anti-inflammatory activity. This fact may also be explained by the different hydrolysis patterns of enzymes selected to hydrolyse AP, leading to a different degree of disruption of pectin chains.

The DM also exerts a great influence on pectin bioactivity. In this sense, in THP-1 macrophages treated with highly methyl-esterified HG branched by arabinogalactans and arabinans, TNF-α and IL1-β secretion was reduced in the presence of a pro-inflammatory agent (de Oliveira et al., 2017). Similar *in vitro* immunostimulatory properties of sweet pepper pectin were also preserved in modified fractions obtained by acid hydrolysis where side chain had been removed showing lower M_W and the DM (do Nascimento et al., 2017). In addition, lavender pectic polysaccharides containing a low-acetylated and high-methoxylated HG domain and RG-I fragments rich in arabinan and arabinogalactan showed anti-inflammatory activity on murine macrophages (Georgiev et al., 2017b), and smaller pectic oligosaccharide chains may also regulate anti-inflammatory cytokine secretion (IL-1 receptor antagonist and IL-10) and may also inhibit the activity of IL-1β (Tan et al., 2018). Selenylation of low methyl-esterified pectin from *Ulmus pumila* L., containing galactan and glucan in its side chains, inhibited nitric oxide production in RAW 264.7 cells resulting in a potential anti-inflammatory activity (Lee et al., 2018).

Similar results were obtained in *in vivo* studies of pectin anti-inflammatory properties. Faecal microbiota transplantation combined with pectin in DSS-induced colitis mice enhanced its positive effects on colonic ratio and DAI and reduced the expression of TNF-α, IL-1β and IL-6 (Hu et al., 2018). A reduction in IL-1β, IL-6, and TNF-α in murine model of endotoxin shock has been also reported for citrus pectin treatment, as well as a decrease in IL-6 secretion from Toll-like receptor-activated RAW264.7 pretreated with citrus pectin. However, when this study was performed using polygalacturonic acid, treatment was not effective, highlighting the importance of neutral sugars present in side chains of pectin (Ishisono et al., 2017). In our study, a similar reduction in pro-inflammatory cytokine expression was observed for AP pectin and APwA, corroborating its immunostimulatory potential. These results agree with those previously reported by Pacheco et al. (2018) where the expression of the inflammatory cytokine panel in DSS-model of mice colitis was reduced after the administration of CP and orange by-products.

It should be noted that pectin may interact with TLR-4 receptor leading to an overall decrease in the expression of TNF-α, IL-1 and IL-6 biomarkers. This mechanism of action was previously suggested by Liu et al. (2010) who evaluated the protective efficacy on intestinal toxicities and carcinogenesis of an apple oligogalactan by targeting LPS/TLR-4/NF-κB pathway in DSS-treated mice. These authors reported a decrease in the expression of pro-inflammatory cytokines. On the other hand, it has been demonstrated that TLR signalling is suppressed by MUC-1 (Ueno et al., 2008), so AP enhances MUC-1 expression leading to a decrease in TLR-4 levels and therefore, low expression of pro-inflammatory cytokines IL-1, IL-6 and TNF-α. Other studies using CP found that low M_w modified fractions reduced inflammation, fibrosis formation in organs and tissues and cancer progression in a mouse model of colitis-associated colon cancer by inhibiting the carbohydrate-binding protein galectin-3 to its ligand. This process may induce apoptosis of cancerous cells (Eliaz, 2016; Li et al., 2012). In our study, AP modification did not enhance its bioactivity although it should be noted that M_w distribution patterns were similar to unmodified pectin.

An *in vivo* study of anti-inflammatory properties of apple pectin showed a decrease of TLR-4 and TNF-α levels in illeal tissue of diet-induced obese rats (Jiang et al., 2016). In other study, apple pectin fractions rich in galactose reduced tumour development in a mice model of colitis-associated colon cancer through the inhibition of galectin-3 (Li et al., 2012). These results agree with those of our study where a significant reduction in

galactose content present in pectin chains led to a dramatic loss of its anti-inflammatory activity. *In vivo* studies dealing with alternative sources of pectin found that RG-I fractions from potato pectin reduce the proliferation of DLD1 and HCT116 colon cancer cells in a dose- and time-dependent manner, as well as ICAM-I expression. The presence of linear GalA segments as well as neutral sugar sidechains enhanced the bioactivity of these extracts (Maxwell et al., 2015). However, we found that higher linearity and lower extent of branching of modified pectin fractions did not result in an increase of their potential bioactivity, although APwA preserved its anti-inflammatory properties. On the other hand, noni fruit polysaccharides, containing RG-I regions with high neutral sugar contents, such as arabinogalactans and arabinans, reduced DAI and promoted the expression of mucosal and tight junction proteins like Occludin and ZO-1 in DSS-induced IBD mice (Jin et al., 2019). Mzoughi et al. (2018) reported that a pectic polysaccharide from *Suaeda fruticosa* rich in arabinose and galactose exerted an anti-inflammatory effect in rats at doses of 100 mg kg^{-1} body weight. Modified pectic polysaccharide from turmeric showing high neutral sugar content, especially galactose and rhamnose, and low M_w (13 kDa) reduced ulcer in rats by decreasing pro-inflammatory factors like TNF–α and galectin-3 levels (Rajagopal et al., 2018). These studies indicate that a high neutral sugar content and degree of branching of pectin enhance its anti-inflammatory potential. However, in our study the galactose content of pectins (Table S6.2) cannot explain differences observed between anti-inflammatory activity of CP and AP; although the latter presents a more ramified structure. These differences can be also due to low DM of AP, achieving the most relevant decreases in pro-inflammatory cytokine expression. Low methyl-esterified pectin from *Opuntia microdasys* cladodes exerted an anti-inflammatory effect in mice and reduced gastric ulcer in rats at doses of 100 mg kg^{-1} body weight (Jouini et al., 2018). Similarly, low methyl-esterified pectic substances isolated from common pondweed, showed anti-endotoxemic activity in mice by decreasing the expression of TNF-α and IL-1β and increasing the levels of anti-inflammatory cytokines such as IL-10 (Popov et al., 2007).

These studies agree with our results, where low methyl-esterified AP showed a higher immunomodulatory capacity than high methyl-esterified CP used as control. Considering the results obtained in these studies, low methyl-esterified pectins have been proposed as good candidates for their anti-inflammatory properties. The immunostimulatory effects in mice of both low and high methyl-esterified CP was compared, showing that low methyl-esterified pectin decreased TNF-α release and

increased production of the anti-inflammatory cytokine IL-10 while high methyl-esterified pectin had no effect on cytokine production (Popov et al., 2013b). Sahasrabudhe et al. (2018) demonstrated that low methyl-esterified pectin inhibits TLR-2 and specifically inhibits the pro-inflammatory TLR-2–TLR-1 pathway by interacting with TLR-2 through electrostatic forces between non-esterified GalA and positive charges on the TLR-2 ectodomain. This mechanism of action may explain differences observed between high and low methyl-esterified pectins obtained in our study.

As explained, a great variability was observed in MUC-1 and TLR-4 levels in those groups treated with modified AP. Differences between groups were very subtle and biochemical profiles were compared through ANN modelling indicating that galactose content significantly contributes to pectin bioactivity. Argüeso et al. (2009) found that modified citrus pectin induces MUC-1. However, this positive effect was reported to be highly dependent on galactose content of pectin whereas APwG yielded higher MUC-1 expression than APwA in our study.

CONCLUSIONS

Potential anti-inflammatory activity of artichoke pectin and modified pectin fractions in dextran sulfate sodium model of mice colitis has been studied. To corroborate differences observed in experimental data and to establish characteristic expression profiles of biochemical parameters for each type of treatment, ANN models were developed. Original artichoke pectin gave rise to a lower expression of pro-inflammatory cytokines TNF-α, IL-1β, IL-6, intercellular adhesion molecule ICAM-I, iNOS enzyme and TLR-4 receptor in a dose-dependent manner. In addition, artichoke pectin administration led to a higher expression of intestinal barrier proteins MUC-1 and Occludin. On the other hand, structural modification of pectin did not enhance its anti-inflammatory effect. Arabinose-free pectin conserved its anti-inflammatory effect while it was observed that galactose content was especially relevant to preserve pectin bioactivity. In general, artichoke pectin showed an enhanced anti-inflammatory potential in ameliorating some IBD symptoms compared to citrus pectin, which could be attributed to its low degree of methyl-esterification. Further studies are needed to completely elucidate the complex mechanism of action of pectin based on its structural features.

General discussion

7. GENERAL DISCUSSION

There is an overall interest in finding new ways of revalorization of industrial by-products in order to achieve sustainable development. One promising alternative involves their use as a source of biologically-active ingredients that may be incorporated in functional foods. In this sense, the industrial processing of artichoke produces a high amount of wastes (external bracts, leaves and stems), which is close to 60% of the harvested artichoke. These by-products are known to be a good source of biologically active ingredients such as inulin, phenolic compounds, and inositols, however as already said, there are scarce studies about extraction of pectin. These are a type of soluble fibre widely used as valuable functional food ingredient due to their technological and biological properties. Moreover, by partial depolymerization of pectins through chemical or enzymatic methods pectic-oligosaccharides (POS) can be obtained and they are one of the most promising candidates to be recognized as prebiotics.

Therefore, in this PhD Thesis a study of different methods to extract pectin from industrial by-products of artichoke as well as POS production by enzymatic depolymerization of artichoke pectin has been carried out. Considering the complexity of pectin structure new methodologies are needed to characterize pectin isolated from novel sources like artichoke, and POS obtained from artichoke pectin hydrolysis, as well as other traditional sources like citrus and apple pectins. With this aim, GC-MS and other relatively simple chromatographic techniques have been used in combination with advanced computational methods such as *in silico* fragmentation and machine learning models. This multidisciplinary approach was developed to assist spectra interpretation of pectic polysachharides and POS in order to gain more information about the structures present in pectin chains and reaction mixtures.

Taking into account these considerations, **Chapter I** of this PhD Thesis (**Articles I** and **II**) aimed at developing an efficient method of pectin extraction from artichoke industrial by-products (external bracts, leaves and stems) using environmentally-friendly techniques such as enzymatic and US-assisted extraction. Then, structural characteristics of pectin extracted by different methods were determined and compared in order to evaluate industrial artichoke by-products as an alternative source of pectin

that may exhibit specific structural features associated with specific biological activities. Optimized enzymatic extraction using commercial cellulase preparation Celluclast 1.5®L (Celluclast) led to a very high yield of pectin (20.3%) at 48 h of enzymatic reaction (**Article I**). Enzymatic extraction is an environmentally friendly process that may lead to higher pectin yields than conventional acid extraction. In this sense, Liew et al. (2016) obtained higher pectin yield from passion fruit peels with Celluclast (9.2%) than with citric acid extraction (7.7%). Similarly, Wikiera et al. (2015a) obtained higher yield with Celluclast (19%) than in sulfuric acid (15%) using apple pomace as raw material while Dominiak et al. (2014), who extracted pectin from lime peel using cellulases, reported higher yields (up to 26%) than those obtained with nitric acid.

This pectin had GalA content of 68.7%, above 65% (limit to be considered as the food additive, E-440) (Morris et al., 2013). These GalA content was lower than those reported by Yuliarti et al. (2015) in pomace pectin extracted with Celluclast (83%), and those obtained by Babbar et al. (2015) in pectin extracted from berry pomace and sugar beet pulp using the same enzyme preparation (75 and 90%, respectively). It can be highlighted the high arabinose content (15.9%) present in artichoke pectin structure which was higher than that of pectin extracted from other sources like citrus and apple by-products (Wikiera et al., 2015a; Gómez et al., 2013; Garna et al., 2004) and similar to pectin from sugar beet pulp (Leijdekkers et al., 2015). The high arabinose content of artichoke pectin could confer to this pectin anticarcinogenic properties (Wikiera et al., 2015a). Other singular property of our pectin was the multimodal M_w distribution similar only to other pectins extracted from alternative sources like mango peel (Huang et al., 2018) and different to monomodal distribution of those conventional pectin sources (citrus and apple) (Fishman et al., 2019; Muñoz-Almagro et al., 2018). Besides, it can be considered a low methoxyl pectin (degree of methyl-esterification 19.5%) similar to other pectins from sunflower and lower than those of apple and citrus pectin previously characterized in our laboratory (72-76%) (Muñoz-Almagro et al., 2018).

Therefore, the use of enzymatic preparation Celluclast is a good method to extract pectin from artichoke by-products and it could be an alternative to traditional sources. This enzymatic method, here optimized, allowed obtaining pectin from artichoke by-products with structural characteristics very different to commercial products obtained from other sources. Low methyl-esterified pectins like artichoke pectin require calcium ions for gelling and may form gels with a low soluble solids content and at a wide pH-range values, widening their applications in dietetic and dairy products. These pectins

are recommended for use in products with lower sugar content (Herbstreith & Fox., 2018). In addition, low methyl-esterified pectin may show an enhanced biological activity compared to high methyl-esterified pectin (Jouini et al., 2018).

The following study was focused on to know the influential factors on artichoke pectin extraction to to develop a faster extraction method. Therefore, a comparative study of enzymatic method, here optimized, with the others usually used to obtain pectin as acids, and alternative US-assisted treatments have been carried out (**Article II**). According to the results, acidic extraction employing sodium citrate and nitric acid led to a less ramified structure of pectin showing high GalA contents (82-90%), high linearity of pectin backbone and low extent of RG-I branching. These results were similar to those of Muñoz-Almagro et al. (2018), who extracted pectin at the same conditions from sunflower by-products (85-90%), and those of Yuliarti et al. (2015), who reported GalA contents of 82% in acid-extracted gold kiwifruit pomace pectin. In contrast, these results were higher than those obtained by Babbar et al. (2015), who obtained GalA contents of 63 and 77% in pectin extracted with nitric acid from berry pomace and sugar beet pulp, respectively. US-process led to low extraction yields (<7.1%) while US-treatments applied in combination with enzyme (Celluclast) achieved the highest yields (13.9%), at the same extraction time (6 h). In general, artichoke pectins extracted by these methods show a higher pectic neutral sugar content (up to 28%) compared to traditional sources like citrus and apple (8-15%) (Wikiera et al., 2015a; Gómez et al., 2013; Garna et al., 2004) and similar to pectin from sugar beet pulp (Leijdekkers et al., 2015).

In order to deepen the knowledge of structure of artichoke pectins obtained using the different methods, several mathematical models were built to find reproducible patterns (**Article II**). With this aim, artificial neural networks (ANN) were computed to classify pectins according to the extraction method considering composition parameters, and FT-IR spectra. In ANN built to study pectin composition parameters, the most influential variables were pectin yield and xylose, glucose and unidentified monosaccharides, the latter being especially relevant in enzymatically extracted pectin. With regard to ANN built to classify FT-IR spectra of extracted pectins, wavenumbers (i. e. 1000 – 1154 cm^{-1}) corresponding to hydroxyl and ether groups were relevant in acid and enzymatic treatments while several wavenumbers corresponding to carboxylic acids (i. e. 1700 cm^{-1}) had a low influence in US treatments, indicating the lower presence of GalA units. These two ANN models, giving complementary information,

corroborated differences observed in experimental data and structural similarities and differences could be established in pectin samples and it could be applied at pectin from unknown origin, although deep structural information would not be obtained. For this purpose, a third ANN model was trained on GC-MS spectra of all pectin monomers (identified with standards and unidentified). To elucidate probable chemical structures of these monomers, we developed a new data analysis strategy (**Figure D1**) based on the combination of machine learning algorithms such as ANN and *competitive fragmentation modelling* (CFM-ID) source code developed by Allen et al. (2015) to interpret ANN model. This software is able to simulate the fragmentation process of organic molecules of several MS-based techniques including GC-EI-MS (Allen et al., 2016). Therefore, it is possible to determine the chemical origin and structure of one *m/z* ion. First, we built a library containing theoretical fragmentation patterns of candidate molecules that could be present among pectin monomers (Atmodjo et al., 2013) and pectin-associated compounds present in artichoke (Lattanzio & van Sumer, 1987). Chemical structures of relevant *m/z* fragments in experimental GC-MS spectra were assigned by comparing to our *in silico* fragmentation library. Considering some of these *m/z* ions might be formed only from some monomers, the probable chemical nature of some of these unidentified compounds was elucidated. Apiose, KDO and several methyl-esterified or acetylated sugars might be present in all artichoke pectin samples extracted with different methods while phenolic compounds like ferulic acid might be present in artichoke pectin extracted with nitric acid. This result agrees with those reported by Liu et al. (2019d), who found higher ferulic acid content in sugar beet pectin obtained with sulfuric acid than in the same pectin obtained with ammonium oxalate.

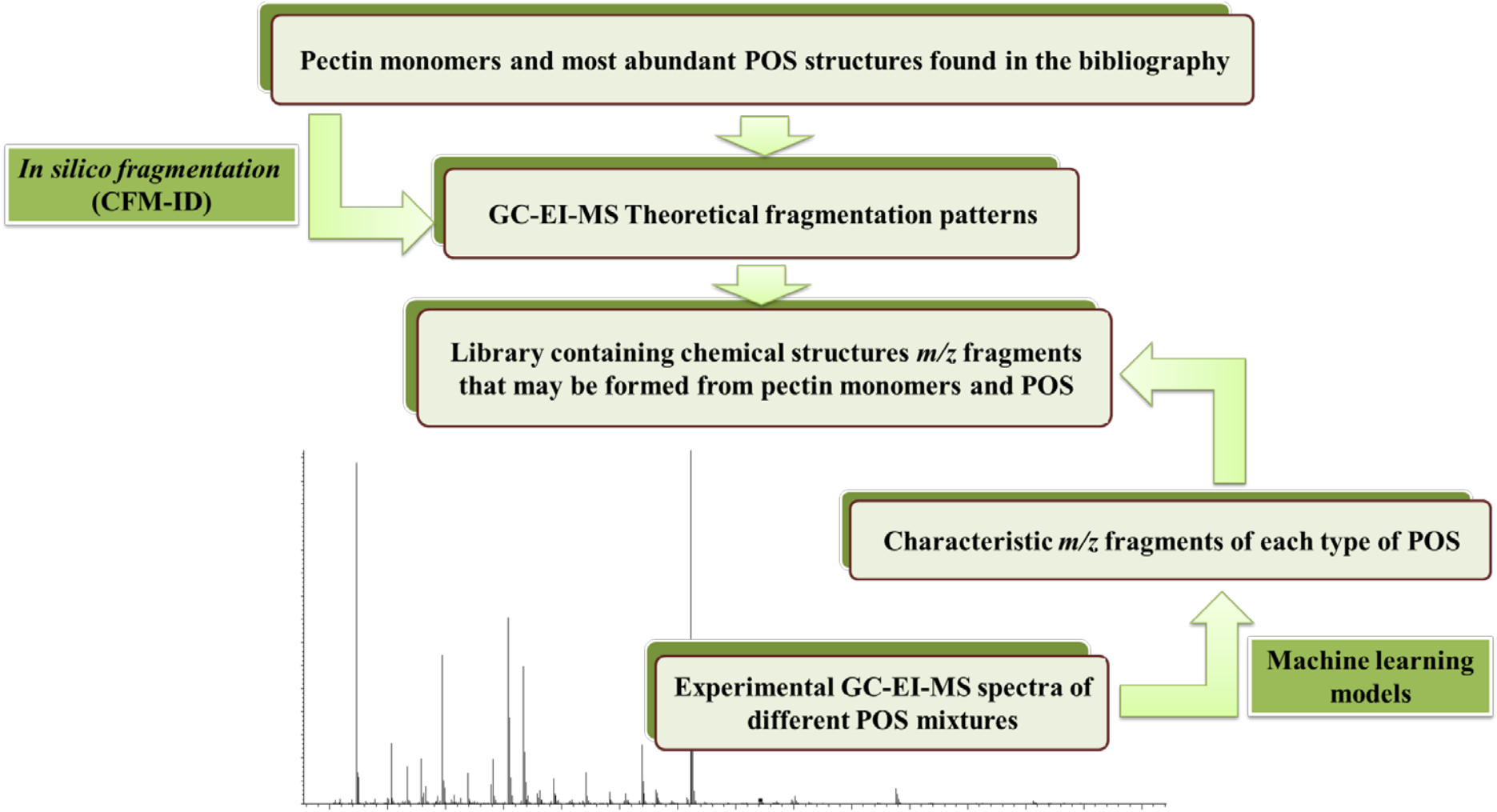


Figure D1. GC-EI-MS ata analysis strategy based on the combination of machine learning models and *in silico* fragmentations. **CFM-ID**: Competitive fragmentation modelling for metabolite identification source code.

In general, several attempts have been made to study and classify carbohydrate structures, using mainly NMR data which gives the most complete structural information (Gerbst et al., 2010). However, NMR spectrometers are expensive and this technique requires, for a complete structural elucidation, that the different compounds under study be isolated. Moreover, high carbohydrate amounts (few mg of pure compounds), which might not be easily obtained due to the pectin complexity, are needed for NMR analysis. We have shown that reproducible structural patterns for several pectic and non-pectic monosaccharides can be established using a simpler technique like GC-MS that could be employed in routine analysis. In addition, targeted elucidation of unknown structures present in complex samples might be possible by this combined approach. Machine learning algorithms like ANN are considered as black-box models which are often difficult to interpret. However, ANN predictions can be associated to chemical structures (i. e. specific fragments of pectic sugars) when combined with *in silico* fragmentation models such as CFM-ID.

Considering POS are one of the most promising candidates to be recognized as prebiotics and their applications as functional ingredients are gaining attention, **Chapter II** of this PhD Thesis is focused on optimization of POS obtainment, including two articles, one corresponding to POS from artichoke pectin obtained in the **Article I** (**Article III**), as well as POS from other sources like citrus and apple pectin and polygalacturonic acid (PGA) (**Article IV**). There is a wide variety of novel pectin sources and enzymes that can be used for POS obtainment. Therefore, it is difficult to predict POS from a specific source. **Articles III** and **IV** determined structural patterns of POS obtained from homogalacturonan domain (PGA), high-methyl-esterified pectin (citrus and apple pectin) and low methyl-esterification pectin (artichoke pectin). Considering the results of both articles, it is possible to tentatively predict some of the most probable POS structures that may be obtained from high or low methyl-esterification pectin using different enzymes. To characterise novel POS formed in reactions, several complementary chromatographic techniques were used.

For POS and modified pectin obtainment from artichoke pectin (**Article III**). different commercial enzyme preparations were selected: Glucanex (β-1,3-glucanase), Pentopan (1,4-endoxylanase), Cellulase from *Aspergillus niger* (Cellulase *An*, cellulase) and Pectinex Ultra Olio (Pectinex, pectin-lyase), all of them showed several secondary activities important to hydrolyse pectin. Cellulase from *Aspergillus niger* and Pectinex Ultra Olio formed large amounts of POS (M_w around 14 and 6.0 kDa) while Glucanex,

at low doses, produced mainly modified pectins (M_w around 60 kDa) and Pentopan scarcely modified the original pectin. It should be noted that HPSEC is not able to determine structural characteristics of individual carbohydrates present in reaction mixtures. When characterising complex POS mixtures, it should be considered that discriminating between neutral and acidic oligosaccharides is especially relevant due to the presence of GalA, the main monomer of pectin. In HPAEC-PAD analysis a wide variety of acid compounds with DP 2, 3 and > 3 were detected corresponding to oligosaccharides with one molecule of GalA and one or more molecules of neutral sugars. Artichoke POS yields were between 33.4 and 311 mg g^{-1}. Pectinex Ultra Olio led to a higher release of GalA, Di-GalA and Tri-GalA while Cellulase from *A. niger* formed significant higher amounts of larger oligosaccharides.

For POS obtainment from citrus and apple pectin and PGA (**Article IV**), Viscozyme (β-glucanase) and Glucanex (at higher concentration than **Article III**) preparations were compared. It was observed that different HPSEC-RID profiles were obtained when different enzyme preparations were applied on the same pectin sources while different pectin sources treated with the same preparation led to similar profiles, in agreement with Olano-Martin et al. (2002). Glucanex produced POS with higher M_w (around 7 kDa) than Viscozyme, (around 2.4 kDa). These results agree with those obtained by other authors so POS mixtures with an average M_w of 3.0 - 3.5 kDa have been obtained from both high and low methyl-esterified pectin using endopolygalacturonases like Viscozyme (one of its secondary activities), that randomly hydrolyses the homogalacturonan backbone. Similarly, a mixture of four endoenzymes (pectin methyl esterase, endo-polygalacturonase (endo-PG), endoarabinosidase, and endogalactosidase) led to POS of around 3 kDa from sugar beet pectin (Ralet et al., 2008). With regard to POS production quantified by HPAEC-PAD, it ranged from 335 to 652 mg g^{-1} pectin for citrus and apple pectin. Viscozyme released higher amounts of GalA, Di-GalA and Tri-GalA. In contrast, Glucanex, produced mainly disaccharides formed of one molecule of GalA attached to one or more neutral monosaccharide.

Our results agree with previous studies that reported the use of Viscozyme to obtain POS from other pectin sources like lemon peel, orange peel, sugar beet pulp, onion skins, rapeseed cake, and PGA (Baldassarre et al., 2018; Elst et al., 2018; Cobs-Rosas et al., 2015; Combo et al., 2012; Gómez et al., 2013; Martinez Sabajanes et al., 2012; Martínez et al., 2009a,b), where high amounts of low M_w POS were obtained and, in

general, it has been observed the influence of enzyme and reaction conditions over POS release and their structural properties (Babbar et al., 2016a).

These different hydrolysis patterns of POS obtained from different sources might be due to diverse activities of these enzymes. In HPAEC-PAD the identification of novel POS formed during enzymatic hydrolysis depends on POS standards that are expensive (Di-GalA and Tri-GalA were used as standards), but many of them are not commercially available. Consequently, HPAEC-PAD cannot elucidate the specific structures of oligosaccharides released, so most POS quantified remain unidentified.

To accurately determine structural features of unknown POS derived from the hydrolysis of all studied pectins (**Articles III** and **IV**), GC-EI-MS can be used as an alternative technique in combination with several data analysis tools: *in silico* fragmentation and machine learning models. First, we added theoretical fragmentation patterns of the most abundant POS structures derived from HG and RG-I to our *in silico* GC-EI-MS fragmentation library. Oligosaccharide GC-EI-MS fragmentation is a complex process that produces intricate data (Vogt et al., 2003), and machine learning models shows two main advantages over other data analysis tools: i) it allows classifying GC-EI-MS spectra of oligosaccharides showing significantly higher accuracies than traditional methods; ii) relationships between specific fragments that could be related to specific hydrolysates can be proposed (**Figure D2**). Association rules between fragments established by machine learning can be calculated to summarize these relationships (**Article IV**). These rules are highly reproducible and could be generalized and applied on similar oligosaccharide structures or POS from other sources. Considering all ions included in one rule are originated together, these ions derive from common larger oligosaccharide structures. So, fragmentation rules calculated allow reconstructing original oligosaccharide molecules based on their fragmentation patterns (**Table D1**). As explained, these POS structures above discussed could not have been elucidated based only in traditional data analysis methods.

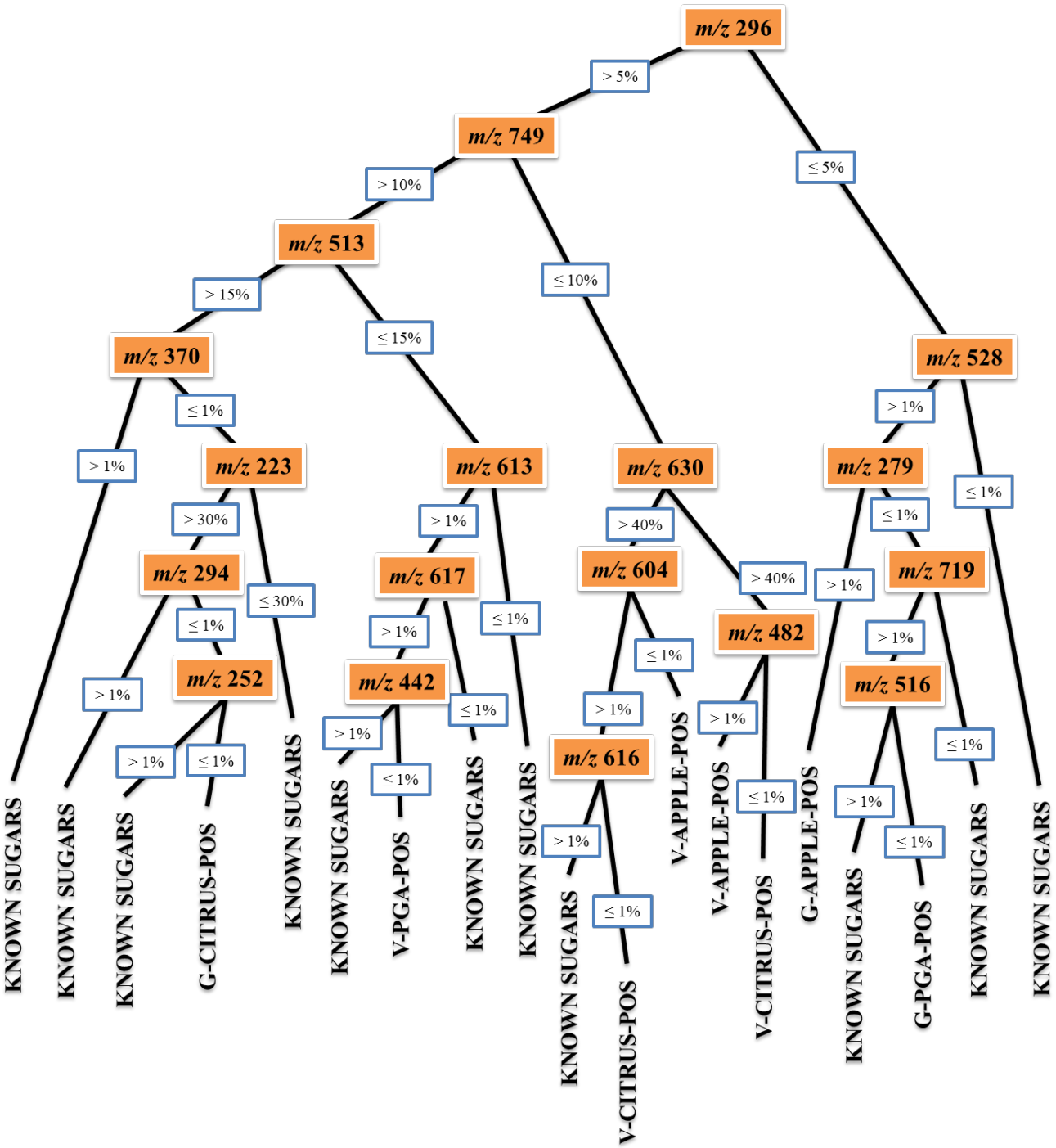


Figure D2. Example of a random forest (RF) tree where a set of classification rules are learnt directly from the input data (GC-EI-MS spectra of different pectic oligosaccharides, POS), considering *m/z* ion abundances (%). **G-**: -: Glucanex preparation, **V-**: Viscozyme preparation.

Table D1. Most probable units found in pectic oligosaccharides (POS) structures produced from polygalacturonic acid (PGA), high methyl-esterification pectins (citrus and apple) and low methyl-esterification pectin (artichoke) by enzymatic hydrolysis using different enzyme preparations (**Articles III** and **IV**). Some larger POS structures containing these units are present or released in functional yogurts containing citrus pectin (**CP**), artichoke pectin (**AP**), citrus POS (**CPOS**) and artichoke POS (**APOS**) (**Article V**). **Xyl**: Xylose, **Ara**: Arabinose, **Rha**: Rhamnose, GalA: Galacturonic acid, **PG**: Polygalacturonase, **exo-Gal**: Exo-β-D-galactofuranosidase, **CAN**: Cellulase from *Aspergillus niger*, **PUO**: Pectinex Ultra Olio.

		POS containing GalA and Rha units			
Substrate	**Enzyme**	**GalA-α(1,2)-Rha**	**GalA-α(1,2)-Rh (acetyl-esterified)**	**Rha-α(1,4)-GalA**	**Rha-α(1,4)-GalA (acetyl-esterified)**
PGA	Viscozyme (endo-PG)	Yes	Yes	-	-
PGA	Glucanex (exo-Gal)	-	-	-	-
Citrus	Viscozyme (endo-PG)	Yes	Yes	-	-
Citrus	Glucanex (exo-Gal)	-	-	-	-
Apple	Viscozyme (endo-PG)	-	Yes	-	-
Apple	Glucanex (exo-Gal)	-	Yes	-	-
Artichoke	Glucanex (exo-Gal)	-	Yes	-	Yes
Artichoke	Pentopan (1,4-endoxylanase)	Yes	Yes	-	Yes
Artichoke	CAN (Cellulase)	-	Yes	-	-
Artichoke	PUO (Pectin-lyase)	Yes	Yes	Yes	Yes
Larger POS structures found in functional yogurts					
Present in yogurts containing CP, CPOS and APOS		DiGalA-α(1,2)-Rha-α(1,4)-Gal			
Present in yogurts containing APOS		GalA-α(1,2)-Rha			
Released from yogurts containing CP, AP		GalA-α(1,2)-Rha			
Released from yogurts containing APOS			GalA-α(1,4)-GalA(Acet)-α(1,2)-Rha		
			DiGalA-α(1,4)-GalA(Acet)-α(1,2)-Rha		
			TriGalA-α(1,4)-GalA(Acet)-α(1,2)-Rha		

Table D1. Cont.

		POS containing GalA and Xy unitsl			POS containing Ara and Rha units
Substrate	**Enzyme**	**Xyl-α(1,3)-GalA**	**Xyl-α(1,3)-GalA (methyl-esterified)**	**Xyl-α(1,3)-GalA (acetyl-esterified)**	**Ara-α(1,4)-Rha**
PGA	Viscozyme (endo-PG)	-	-	Yes	-
PGA	Glucanex (exo-Gal)	Yes	Yes	Yes	-
Citrus	Viscozyme (endo-PG)	-	-	Yes	-
Citrus	Glucanex (exo-Gal)	-	-	-	-
Apple	Viscozyme (endo-PG)	Yes	Yes	Yes	-
Apple	Glucanex (exo-Gal)	-	-	-	-
Artichoke	Glucanex (exo-Gal)	-	-	Yes	-
Artichoke	Pentopan (1,4-endoxylanase)	Yes	Yes	Yes	Yes
Artichoke	CAN (Cellulase)	-	-	Yes	-
Artichoke	PUO (Pectin-lyase)	Yes	Yes	Yes	-

Larger POS structures found in functional yogurts		
Present in yogurts containing APOS	Xyl-α(1,3)-GalA	
Present in yogurts containing CP, CPOS and APOS	Xyl-α(1,3)-TriGalA	
Released from yogurts containing CP and AP	Xyl-α(1,3)-GalA	
	Xyl-α(1,3)-DiGalA	
Released from yogurts containing APOS	Xyl-α(1,3)-TetraGalA	Xyl-α(1,3)-GalA(Acet)-α(1,4)-GalA
		Xyl-α(1,3)-GalA(Acet)-α(1,4)-DiGalA

Given the multidisciplinary approach of this PhD Thesis, the third chapter dealt with practical applications of previously extracted and characterized substrates. With this aim, functional yogurts containing artichoke pectin and POS were elaborated (**Article V**). We select this food because yogurt is one of the most popular fermented milk consumed worldwide that improves the balance of microbiota and is a good vehicle to delivery functional ingredients. In this work behaviour of carbohydrate fractions of functional yogurt containing artichoke and citrus pectin and POS during elaboration has been studied. For this purpose, a comprehensive characterization of the carbohydrate fraction of each type of functional yogurt was performed using GC-EI-MS and MALDI-TOF-MS as complementary techniques (**Article V**). Cell count, pH, lactose content and lactic and acetic acid production accomplished the quality standards of yogurts. Other yogurt sugars from pectin (arabinose, GalA, POS) and milk (glucose, galactose, *myo*-inositol and GOS) were determined. GOS were formed during yogurt manufacture in agreement with Toba et al. (1982) and Martínez-Villaluenga et al. (2008). Interestingly, small amounts of several POS were also released during fermentation step and their structures were elucidated following the GC-EI-MS data analysis strategy developed in this PhD Thesis.

It has been demonstrated that machine learning models allow establishing structural differences between various types of carbohydrates (yogurt sugars) as well as POS from different sources (citrus and artichoke pectin) (**Figure D3**). Artichoke POS present in yogurts showed a wide variety of fragments from structures formed by xylose and rhamnose molecules attached to one or more GalA residues.

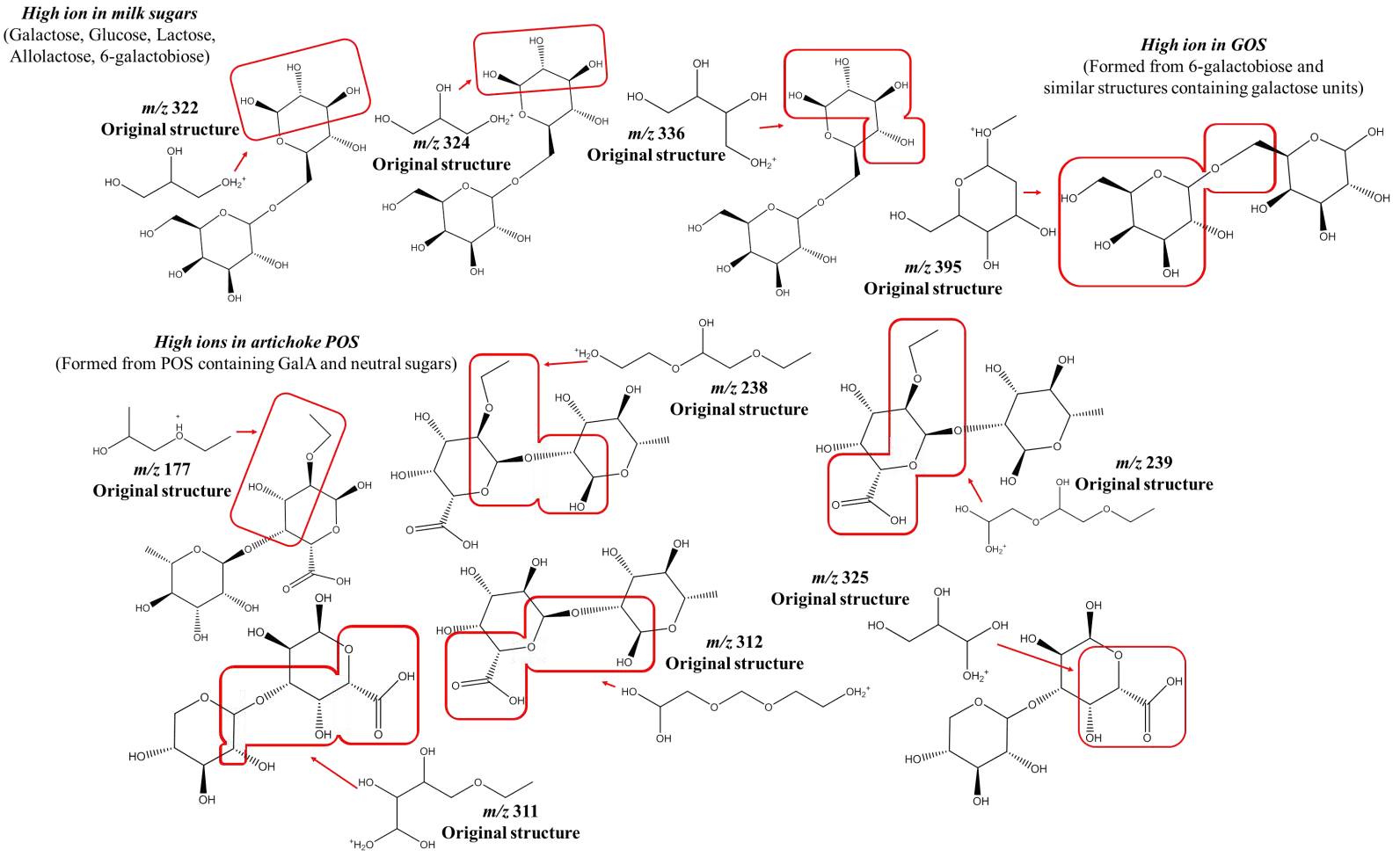


Figure D3. Chemical origin of relevant *m/z* ions of carbohydrates found in yogurts supplemented with citrus and artichoke pectins and POS. Only one *m/ z* ion posible conformation is shown.

In order to gain more information from GC-EI-MS spectra, quantitative structure-retention time relationships were calculated considering the fragmentation patterns above described. These relationships provide additional information about the abundance of specific fragments on larger oligosaccharide structures and their influence on the relative retention time (RRT) of each molecule. Therefore, some specific ions (i. e. *m/z* 239, 340, 442) derived from Rha-α(1,4)-GalA, GalA-α(1,2)-Rha (acetylated or not) or Xyl-α(1,3)-GalA (acetylated) had a positive influence on the RRT, indicating the presence of this substructures in larger POS molecules (tri- and tetrasaccharides) (**Table D1**). These structures were released during fermentation. Then, to correlate all data generated, the abundances of *m/z* ions attributed to specific carbohydrate structures found in specific types of functional yogurts were related to their RRT by a convolutional neural network (CNN). This model established structural patterns to compare the complete carbohydrate fraction of functional yogurts and not only for individual GC-EI-MS spectra. CNN confirmed that tri- and tetra-POS formed by Rha/Xyl attached to acetylated GalA units were specifically released during fermentation of artichoke POS-supplemented yogurts. The presence of these structures was also corroborated by MALDI-TOF-MS analysis, highlighting the high predictive power of machine learning models for structure elucidation based solely on spectral data from a relatively simple technique like GC-MS. In general, this model could be applied to automatically determine structural patterns in the carbohydrate fraction of any kind of food matrices.

Finally, biological activity of artichoke pectin extracted from by-products was evaluated (**Article VI**) and compared to citrus pectin. It should be noted that citrus pectin shows a high degree of methyl-esterification, and this property exerts a great influence on the biological activity of pectin. Anti-inflammatory properties of artichoke pectin and modified fractions (galactose- and arabinose-free) were assessed in dextran sulphate sodium (DSS) model of mice colitis to determine the influence of pectic neutral sugar content.. Artichoke pectin administered at doses of 80 mg kg^{-1} ameliorated the symptoms of Inflammatory Bowel Disease (IBD) considering the macroscopic indicators and cytokine expression profiles. Differences in biochemical profiles obtained by each fraction were sometimes not significant, so they were compared through ANN modelling, indicating that galactose content was the main factor contributing to pectin bioactivity, while effect of arabinose presence was lower. Original artichoke pectin treatment reduced the expression of pro-inflammatory

cytokines like IL-6 and TNF-α and increased the expression of intestinal barrier proteins such as Occludin and MUC-1. Arabinose free- artichoke pectin kept anti-inflammatory properties although these beneficial effects were lower than those of original artichoke pectin. This pectin showed a higher anti-inflammatory potential compared to citrus pectin due to its high neutral sugar content, in accordance to various studies (Jin et al., 2019; Hu et al., 2018; Pacheco et al., 2018; Ishisono et al., 2017). Galactose content of pectin has proven to be especially relevant through the inhibition of carbohydrate-binding protein galectin-3 (Li et al., 2012) which induces apoptosis of cancerous cells (Eliaz, 2017). We have demonstrated that a significant reduction in galactose content present in pectin chains led to a dramatic loss of its anti-inflammatory activity. On the other hand, the degree of methyl-esterification is another structural feature that may exert a great influence on the anti-inflammatory activity of pectin. Thus, artichoke pectin which is low methoxylated showed a higher immunomodulatory capacity than high methyl-esterified citrus pectin used as control in agreement with the results reported by several authors (Jouini et al., 2018; Sahasrabudhe et al. 2018). In our study, degree of methyl-esterification of pectin demonstrated to be more relevant than galactose content considering the high galactose contents of citrus pectin used as control. It has been demonstrated that low methoxylated pectin inhibits with pro-inflammatory toll-like receptors through electrostatic forces between non-esterified GalA and positive charges on the TLR2 ectodomain (Sahasrabudhe et al., 2018). Low methyl-esterified artichoke pectin enhances MUC1 expression leading to a decrease in TLR4 levels and therefore, low expression of pro-inflammatory cytokines IL-1, IL-6 and TNF-α. Therefore, this pectin could be proposed as good candidates for their anti-inflammatory properties and enzymatic modification of this pectin did not enhance its anti-inflammatory effect.

This thesis describes pectin extraction and characterization from a novel source (artichoke industrial by-products) showing different structural features compared to traditional sources of pectin. In addition, enzymatic hydrolysis of pectin has been studied to obtain a wide number of POS structures. Practical applications of these products have been evaluated through their incorporation in functional yogurt formulation as well as the assessment of the anti-inflammatory activity of artichoke pectin *in vivo*. To elucidate the complex structures of pectin and POS mixtures and to correlate their compositional features with biological data, an advanced data analysis strategy based on different computational models (machine learning and *in silico*

fragmentation) has been developed. These models allow discriminating between different carbohydrate structures and conformations without the need of any purification step. This fact is an advantage when characterising complex structures present in reaction mixtures in order to evaluate their potential bioactivity. In addition, these models allow determining differences in complex profiles of biochemical marker determined in biological samples.

At last, in future trends, machine learning could be employed for an *in silico* screening of the most relevant biological activities of a given molecule, based on structural similarities to known drugs or bioactive compounds. The ability to determine structural similarities and differences among potentially prebiotic carbohydrates can be of great importance to establish structure-function relationships which could be very useful to extrapolate results from biological activities, previous to *in vitro* and *in vivo* assays.

Conclusions

8. CONCLUSIONS

1. The optimized method of pectin extraction from artichoke by-products using a cellulase gave rise a pectin with high yield and galacturonic acid content (> 65%) therefore it could be considered as a food additive (E 440). This pectin shows characteristic structural properties including its low degree of methyl-esterification, multimodal molecular weight distribution and high neutral sugar contents, especially arabinose.

2. A combined method based on the utilization of cellulase enzymes and ultrasounds could be used for the obtainment of artichoke pectin in short times, with good yields and without important molecular modification.

3. GC-EI-MS patterns of artichoke pectin structures extracted by different methods have been elucidated indicating the possible presence of several neutral sugars, methyl-esterified/acetylated acids and phenolic compounds.

4. Pectic oligosaccharides (POS) mixtures have been obtained from artichoke pectin using commercial enzymes with different activities. Among studied preparations, Cellulase from *Aspergillus niger* and Pectinex Ultra Olio formed large amounts of POS (M_w around 14 and 6.0 kDa).

5. Enzymatic depolymerization of two common highly methyl-esterified pectins (citrus and apple) has been studied and compared to artichoke POS patterns. In these substrates, Viscozyme preparation produced mainly POS with an M_w of 2.5 kDa while Glucanex led to several POS fragments (M_w around of 7−11 kDa and 2.6−3.8 kDa).

6. An advanced data analysis strategy based on the combination of machine learning algorithms and *in silico* fragmentation has been developed to elucidate structural patterns from GC-EI-MS spectra of novel oligosaccharides like POS when no reference standard or spectra library is available.

7. Functional yogurts containing citrus and artichoke pectin and POS have been elaborated. A wide variety of POS structures were released during their manufacture in small amounts. These structures were elucidated by GC-EI-MS in combination with machine learning models and then confirmed by MALDI-TOF-MS.

8. Artichoke pectin reduced pro-inflammatory cytokine expression *in vivo* in DSS mice model of colitis showing an enhanced bioactivity compared to high methyl-esterified citrus pectin. Expression profiles were established through artificial neural networks and it was observed that galactose content was especially relevant to preserve anti-inflammatory activity of pectin.

Considering the individual conclusions from each chapter, the general conclusion of this PhD Thesis is:

Artichoke pectin and POS have been obtained using enzymatic methods. A comprehensive characterization of pectin chains and POS structures has been carried out by a novel computational approach. Artichoke pectin and POS may be incorporated in food matrices and some of these products exerted and enhanced anti-inflammatory activity *in vivo* compared to pectin obtained from conventional sources.

Conclusiones

9. CONCLUSIONES

1. El método optimizado de extracción de pectina a partir de subproductos de alcachofa utilizando una celulasa dio lugar a altos rendimientos de pectina con un elevado contenido en ácido galacturónico (> 65%), por lo que podría utilizarse como aditivo alimentario (E-440). Esta pectina presentó una serie de características como su bajo grado de metil-esterificación, su distribución multimodal de masas moleculares y su alto contenido en azúcares neutros, especialmente arabinosa.

2. El método combinado basado en la utilización de una enzima celulasa y ultrasonidos podría ser utilizado para la obtención de pectina de alcachofa a tiempos cortos, mostrando altos rendimientos y sin una modificación molecular importante.

3. Se han elucidado patrones estructurales de las pectinas de alcachofa extraídas por diferentes métodos a través de sus patrones de GC-EI-MS. Estos resultados indican la presencia de diversos azúcares neutros, ácidos metil- y acetyl-esterificados y compuestos fenólicos.

4. Se han obtenido mezclas de oligosacáridos pécticos (POS) a partir de pectina de alcachofa utilizando enzimas comerciales con diferentes actividades. De entre todos los preparados estudiados, la Celulasa de *Aspergillus niger* y Pectinex Ultra Olio dieron lugar a un elevado rendimiento de POS (con una masa molecular media de 14 y 6,0 kDa).

5. Se ha estudiado la despolimerización enzimática de dos pectinas comunes de alto grado de metil-esterificación (cítricos y manzana), comparándose con los patrones de POS de alcachofa obtenidos. El preparado Viscozyme formó principalmente POS con una masa molecular de 2,5 kDa mientras que la enzima Glucanex dio lugar a diversas cadenas de POS (con masas moleculares en torno a 7-11 kDa y 2,6-3,8 kDa).

6. Se ha desarrollado una estrategia de análisis de datos avanzada, basada en la combinación de los algoritmos de aprendizaje automático y la fragmentación *in silico*, para estudiar los patrones de fragmentación de GC-EI-MS de nuevas mezclas de oligosacáridos como los POS cuando no se dispone de patrones o librerías de referencia.

7. Se han elaborado yogures funcionales con pectina y POS de cítricos y alcachofa. Una gran variedad de estructuras de POS se liberaron en pequeñas cantidades durante su elaboración. Estas estructuras se caracterizaron mediante GC-EI-MS en combinación con los modelos de aprendizaje automático, y posteriormente se verificaron mediante MALDI-TOF-MS.

8. La pectina de alcachofa redujo la expresión de citoquinas pro-inflamatorias *in vivo* en un modelo de colitis inducida por DSS en ratones, mostrando una bioactividad mejorada con respecto a la pectina de cítricos de alto grado de metil-esterificación. Se han establecido perfiles de expresión mediante redes neuronales artificiales y se observó que el contenido en galactosa resultó ser especialmente importante para mantener la actividad anti-inflamatoria de la pectina.

Teniendo en cuenta las conclusiones individuales de cada capítulo, la conclusión general de esta Tesis Doctoral es:

Se han obtenido pectina y POS de alcachofa utilizando métodos enzimáticos. Se ha realizado una caracterización exhaustiva de las estructuras de pectinas y POS mediante un método computacional novedoso. La pectina y los POS de alcachofa podrían ser incorporados en matrices alimentarias y algunos de estos productos ejercen una actividad anti-inflamatoria *in vivo* mejorada con respecto a pectinas obtenidas de fuentes convencionales.

References

10. REFERENCES

Abboud, K. Y., da Luz, B. B., Dallazen, J. L., de Paula Werner, M. F., Cazarin, C. B. B., Junior, M. R. M., Iacomini, M., & Cordeiro, L. M. (2019). Gastroprotective effect of soluble dietary fibres from yellow passion fruit (Passiflora edulis f. flavicarpa) peel against ethanol-induced ulcer in rats. *Journal of functional foods*, *54*, 552-558.

Acharjee, A., Prentice, P., Acerini, C., Smith, J., Hughes, I. A., Ong, K., Griffin, J. L.; Dunger, D., & Koulman, A. (2017). The translation of lipid profiles to nutritional biomarkers in the study of infant metabolism. *Metabolomics*, *13*(3), 25.

Acunha, T., García-Cañas, V., Valdés, A., Cifuentes, A., & Simó, C. (2018). Metabolomics study of early metabolic changes in hepatic HepaRG cells in response to rosemary diterpenes exposure. *Analytica chimica acta*, *1037*, 140-151.

Adetunji, L. R., Adekunle, A., Orsat, V., & Raghavan, V. (2017). Advances in the pectin production process using novel extraction techniques: A review. *Food Hydrocolloids*, *62*, 239-250.

Albelwi, S., & Mahmood, A. (2017). A framework for designing the architectures of deep convolutional neural networks. *Entropy*, *19*(6), 242.

Aldrich, E. (2013). Wavelets: A package of functions for computing wavelet filters, wavelet transforms and multiresolution analyses. *R package version 0.3-0*. URL https://CRAN.R-project.org/package=wavelets. Last accessed: 24/06/2019

Allaire, J. J., Chollet, F. (2018). keras: R Interface to 'Keras'. *R package version 2.2.4*. https://CRAN.R-project.org/package=keras. Last accessed: 02/07/2019

Allard, P. M., Péresse, T., Bisson, J., Gindro, K., Marcourt, L., Pham, V. C., Roussi, F., Litaudon, M., & Wolfender, J. L. (2016). Integration of molecular networking and in-silico MS/MS fragmentation for natural products dereplication. *Analytical chemistry*, *88*(6), 3317-3323.

Allen, F., Pon, A., Greiner, R., & Wishart, D. (2016). Computational prediction of electron ionization mass spectra to assist in GC/MS compound identification. *Analytical chemistry*, *88*(15), 7689-7697.

Allen, F., Greiner, R., & Wishart, D. (2015). Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification. *Metabolomics*, *11*(1), 98-110.

Alpaydin, E. (2014). *Introduction to machine learning*. MIT press.

Al-Tamimi, M. A. H. M., Palframan, R. J., Cooper, J. M., Gibson, G. R., & Rastall, R. A. (2006). In vitro fermentation of sugar beet arabinan and arabino-oligosaccharides by the human gut microflora. *Journal of Applied Microbiology*, *100*(2), 407-414.

Álvarez, B. G. (2015). *Obtención, purificación, caracterización y evaluación de nuevos* prebióticos *a partir de subproductos agroindustriales* (Doctoral dissertation, Universidade de Vigo).

Alygizakis, N. A., Gago-Ferrero, P., Hollender, J., & Thomaidis, N. S. (2019). Untargeted time-pattern analysis of LC-HRMS data to detect spills and compounds with high fluctuation in influent wastewater. *Journal of hazardous materials*, *361*, 19-29.

Argüeso, P., Guzman-Aranguez, A., Mantelli, F., Cao, Z., Ricciuto, J., & Panjwani, N. (2009). Association of cell surface mucins with galectin-3 contributes to the

ocular surface epithelial barrier. *Journal of Biological Chemistry, 284*(34), 23037-23045.

Arioui, F., Ait Saada, D., & Cheriguene, A. (2017). Physicochemical and sensory quality of yogurt incorporated with pectin from peel of Citrus sinensis. *Food science & nutrition, 5*(2), 358-364.

Aruwa, C. E., Amoo, S. O., & Kudanga, T. (2018). Opuntia (Cactaceae) plant compounds, biological activities and prospects–A comprehensive review. *Food Research International*, *112*, 328-344.

Association of Official Analytical Chemists (AOAC). (1990a). Method 920.151. In Association of Official Analytical Chemists (Eds.), Official methods of analysis of the AOAC (15th ed.). Arlington, VA: EEUU.

Association of Official Analytical Chemists (AOAC). (1990b). Method 920.152. In Association of Official Analytical Chemists (Eds.), Official methods of analysis of the AOAC (15th ed.). Arlington, VA: EEUU.

Association of Official Analytical Chemists (AOAC). (1990c). Method 920.39. In Association of Official Analytical Chemists (Eds.), Official methods of analysis of the AOAC (15th ed.). Arlington, VA: EEUU.

Atmodjo, M. A., Hao, Z., & Mohnen, D. (2013). Evolving views of pectin biosynthesis. *Annual review of plant biology, 64*.

Babbar, N., Dejonghe, W., Sforza, S., & Elst, K. (2017). Enzymatic pectic oligosaccharides (POS) production from sugar beet pulp using response surface methodology. *Journal of food science and technology, 54*(11), 3707-3715.

Babbar, N., Baldassarre, S., Maesen, M., Prandi, B., Dejonghe, W., Sforza, S., & Elst, K. (2016a). Enzymatic production of pectic oligosaccharides from onion skins. *Carbohydrate polymers*, *146*, 245-252.

Babbar, N., Dejonghe, W., Gatti, M., Sforza, S., & Elst, K. (2016b). Pectic oligosaccharides from agricultural by-products: production, characterization and health benefits. *Critical reviews in biotechnology*, *36*, 594-606.

Babbar, N., Roy, S. V., Wijnants, M., Dejonghe, W., Caligiani, A., Sforza, S., & Elst, K. (2015). Effect of extraction conditions on the saccharide (neutral and acidic) composition of the crude pectic extract from various agro-industrial residues. *Journal of agricultural and food chemistry*, *64*(1), 268-276.

Baldassarre, S., Babbar, N., Van Roy, S., Dejonghe, W., Maesen, M., Sforza, S., & Elst, K. (2018). Continuous production of pectic oligosaccharides from onion skins with an enzyme membrane reactor. *Food chemistry*, *267*, 101-110.

Banerjee, J., Singh, R., Vijayaraghavan, R., MacFarlane, D., Patti, A. F., & Arora, A. (2017). Bioactives from fruit processing wastes: Green approaches to valuable chemicals. *Food chemistry*, *225*, 10-22.

Behkami, S., Zain, S. M., Gholami, M., & Khir, M. F. A. (2019). Classification of cow milk using artificial neural network developed from the spectral data of single- and three-detector spectrophotometers. *Food chemistry*, *294*, 309-315.

Behrmann, J., Etmann, C., Boskamp, T., Casadonte, R., Kriegsmann, J., & Maaβ, P. (2017). Deep learning for tumor classification in imaging mass spectrometry. *Bioinformatics*, *34*(7), 1215-1223.

Belščak-Cvitanović, A., Đorđević, V., Karlović, S., Pavlović, V., Komes, D., Ježek, D., & Nedović, V. (2015). Protein-reinforced and chitosan-pectin coated alginate microparticles for delivery of flavan-3-ol antioxidants and caffeine from green tea extract. *Food hydrocolloids*, *51*, 361-374.

Bergmeir, C., & Benitez J. M. (2012). Neural Networks in R Using the Stuttgart Neural Network Simulator: RSNNS. *Journal of Statistical Software*, 46, 1-26. URL http://www.jstatsoft.org/v46/i07/. Last accessed: 24/06/2019

Berlin, A., Maximenko, V., Gilkes, N., & Saddler, J. (2007). Optimization of enzyme complexes for lignocellulose hydrolysis. *Biotechnology and bioengineering*, *97*(2), 287-296.

Bitchagno, G. T., & Tanemossu, S. A. F. (2019). Computational methods for NMR and MS for structure elucidation III: more advanced approaches. *Physical Sciences Reviews*.

Blaženović, I., Kind, T., Ji, J., & Fiehn, O. (2018). Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. *Metabolites*, *8*(2), 31.

Boccard, J., & Rudaz, S. (2014). Harnessing the complexity of metabolomic data with chemometrics. *Journal of Chemometrics*, *28*(1), 1-9.

Böcker, S., & Rasche, F. (2008). Towards de novo identification of metabolites by analyzing tandem mass spectra. *Bioinformatics*, *24*(16), i49-i55.

Bonnin, E., Garnier, C., & Ralet, M. C. (2014). Pectin-modifying enzymes and pectin-derived materials: applications and impacts. *Applied microbiology and biotechnology*, *98*(2), 519-532.

Bonnin, E., Dolo, E., Le Goff, A., & Thibault, J. F. (2002). Characterization of pectin subunits released by an optimized combination of enzymes. *Carbohydrate Research*, *337*(18), 1687-1696.

Brobst, K. M., & Lott, C. E. (1966). Gas-Chromatographic Investigation of Hydroxyethyl Amylose Hydrolyzates. *Analytical Chemistry*, *38*(12), 1767-1770.

Brosseau, C., Selle, A., Palmer, D. J., Prescott, S. L., Barbarot, S., & Bodinier, M. (2019). Prebiotics: Mechanisms and Preventive Effects in Allergy. *Nutrients*, *11*(8), 1841.

Broxterman, S. E., Picouet, P., & Schols, H. A. (2017). Acetylated pectins in raw and heat processed carrots. *Carbohydrate polymers*, *177*, 58-66.

Bruzantin, F. P., Daniel, J. L. P., Da Silva, P. P. M., & Spoto, M. H. F. (2016). Physicochemical and sensory characteristics of fat-free goat milk yogurt with added stabilizers and skim milk powder fortification. *Journal of dairy science*, *99*(5), 3316-3324.

Caffall, K. H., & Mohnen, D. (2009). The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydrate research*, *344*(14), 1879-1900.

Canteri, M. H. G., Nogueira, A., de Oliveira Petkowicz, C. L., & Wosiacki, G. (2012). Characterization of Apple Pectin–A Chromatographic Approach. *Chromatography–The most versatile method of chemical analysis*, 325.

Cardelle-Cobas, A., Martínez-Villaluenga, C., Villamiel, M., Olano, A., & Corzo, N. (2008). Synthesis of oligosaccharides derived from lactulose and Pectinex Ultra SP-L. *Journal of agricultural and food chemistry*, *56*(9), 3328-3333.

Chan, S. Y., Choo, W. S., Young, D. J., & Loh, X. J. (2017). Pectin as a rheology modifier: Origin, structure, commercial production and rheology. *Carbohydrate polymers*, *161*, 118-139.

Chandrapala, J., Oliver, C. M., Kentish, S., & Ashokkumar, M. (2013). Use of power ultrasound to improve extraction and modify phase transitions in food processing. *Food Reviews International*, *29*(1), 67-91.

Chen, L., Liu, J., Zhang, Y., Dai, B., An, Y., & Yu, L. (2015). Structural, thermal, and anti-inflammatory properties of a novel pectic polysaccharide from alfalfa

(Medicago sativa L.) stem. *Journal of agricultural and food chemistry*, *63*(12), 3219-3228.

Cheng, Y., & Bélanger, R. R. (2000). Protoplast preparation and regeneration from spores of the biocontrol fungus Pseudozyma flocculosa. *FEMS microbiology letters*, *190*(2), 287-291.

Cobs-Rosas, M., Concha-Olmos, J., Weinstein-Oppenheimer, C., & Zuniga-Hansen, M. E. (2015). Assessment of antiproliferative activity of pectic substances obtained by different extraction methods from rapeseed cake on cancer cell lines. *Carbohydrate polymers*, *117*, 923-932.

Colantuono, A., Ferracane, R., & Vitaglione, P. (2018). Potential bioaccessibility and functionality of polyphenols and cynaropicrin from breads enriched with artichoke stem. *Food chemistry*, *245*, 838-844.

Combo, A. M. M., Aguedo, M., Quiévy, N., Danthine, S., Goffin, D., Jacquet, N., Blecker, C., Devaux, J., & Paquot, M. (2013). Characterization of sugar beet pectic-derived oligosaccharides obtained by enzymatic hydrolysis. *International journal of biological macromolecules*, *52*, 148-156.

Combo, A. M. M., Aguedo, M., Goffin, D., Wathelet, B., & Paquot, M. (2012). Enzymatic production of pectic oligosaccharides from polygalacturonic acid with commercial pectinase preparations. *Food and bioproducts processing*, *90*(3), 588-596.

Corzo, N., Alonso, J. L., Azpiroz, F., Calvo, M. A., Cirici, M., Leis, R., Lombó, F., Mateos-Aparicio, I., Plou, F. J., Ruas-Madiedo, P., Rúperez, P., Redondo-Cuenca, A., Sanz, M. L., & Celemente, A. (2015). Prebióticos; concepto, propiedades y efectos beneficiosos. *Nutrición Hospitalaria*, *31*(1), 99-118.

de Moura, F. A., Macagnan, F. T., & da Silva, L. P. (2015). Oligosaccharide production by hydrolysis of polysaccharides: a review. *International Journal of Food Science & Technology*, *50*(2), 275-281.

de Oliveira, A. F., do Nascimento, G. E., Iacomini, M., Cordeiro, L. M. C., & Cipriani, T. R. (2017). Chemical structure and anti-inflammatory effect of polysaccharides obtained from infusion of Sedum dendroideum leaves. *International journal of biological macromolecules*, *105*, 940-946.

de Santana, F. B., Neto, W. B., & Poppi, R. J. (2019). Random forest as one-class classifier and infrared spectroscopy for food adulteration detection. *Food chemistry*, *293*, 323-332.

de Vries, R. P., & Visser, J. A. A. P. (2001). Aspergillus enzymes involved in degradation of plant cell wall polysaccharides. *Microbiology and molecular biology reviews*, *65*(4), 497-522.

Delattre, C., Michaud, P., & Vijayalakshmi, M. A. (2008). New monolithic enzymatic micro-reactor for the fast production and purification of oligogalacturonides. *Journal of Chromatography B*, *861*(2), 203-208.

Delgado-Fernández, P., Corzo, N., Olano, A., Hernández-Hernández, O., & Moreno, F. J. (2019a). Effect of selected prebiotics on the growth of lactic acid bacteria and physicochemical properties of yoghurts. *International Dairy Journal*, *89*, 77-85.

Delgado-Fernández, P., Corzo, N., Lizasoain, S., Olano, A., & Moreno, F. J. (2019b). Fermentative properties of starter culture during manufacture of kefir with new prebiotics derived from lactulose. *International Dairy Journal*, *93*, 22-29.

Delgado-Povedano, M. M., & De Castro, M. L. (2015). A review on enzyme and ultrasound: A controversial but fruitful relationship. *Analytica chimica acta*, *889*, 1-21.

Delzenne, N. M., & Roberfroid, M. R. (1994). Physiological effects of non-digestible oligosaccharides. *LWT-Food Science and Technology*, *27*(1), 1-6.

Demirkol, M., & Tarakci, Z. (2018). Effect of grape (Vitis labrusca L.) pomace dried by different methods on physicochemical, microbiological and bioactive properties of yoghurt. *LWT*, *97*, 770-777.

Dias, M. I., Barros, L., Barreira, J. C., Alves, M. J., Barracosa, P., & Ferreira, I. C. (2018). Phenolic profile and bioactivity of cardoon (Cynara cardunculus L.) inflorescence parts: Selecting the best genotype for food applications. *Food chemistry*, *268*, 196-202.

Diplock, A. T. (1999). Scientific Concepts of Functional Foods in Europe–Consensus Document British Journal of Nutrition 81. *S1–S27*.

do Nascimento, G. E., Winnischofer, S. M. B., Ramirez, M. I., Iacomini, M., & Cordeiro, L. M. C. (2017). The influence of sweet pepper pectin structural characteristics on cytokine secretion by THP-1 macrophages. *Food research international*, *102*, 588-594.

Doco, T., Williams, P., Meudec, E., Cheynier, V., & Sommerer, N. (2015). Complex carbohydrates of red wine: characterization of the extreme diversity of neutral oligosaccharides by ESI-MS. *Journal of agricultural and food chemistry*, *63*(2), 671-682.

Dominiak, M., Søndergaard, K. M., Wichmann, J., Vidal-Melgosa, S., Willats, W. G., Meyer, A. S., & Mikkelsen, J. D. (2014). Application of enzymes for efficient extraction, modification, and development of functional properties of lime pectin. *Food hydrocolloids*, *40*, 273-282.

dos Santos, E. K. R., Azoubel, P. M., & Gouveia, E. R. (2017). Better Pectin Yield From Passion Fruit Peel (Passiflora edulis f. flavicarpa): From Shaker or Ultrasound? A Comparison. *Waste and biomass valorization*, *8*(3), 905-910.

Dranca, F., & Oroian, M. (2018). Extraction, purification and characterization of pectin from alternative sources with potential technological applications. *Food Research International*, *113*, 327-350.

Ebringerová, A., & Hromádková, Z. (2010). An overview on the application of ultrasound in extraction, separation and purification of plant polysaccharides. *Central European Journal of Chemistry*, *8*(2), 243-257.

Eliaz, I. (2016). U.S. Patent No. 9,427,449. Washington, DC: U.S. Patent and Trademark Office. URL https://patents.google.com/patent/US9427449B2/en. Last accessed: accessed 12/09/2019

Elst, K., Babbar, N., Van Roy, S., Baldassarre, S., Dejonghe, W., Maesen, M., & Sforza, S. (2018). Continuous production of pectic oligosaccharides from sugar beet pulp in a cross flow continuous enzyme membrane reactor. *Bioprocess and biosystems engineering*, *41*(11), 1717-1729.

Enam, F., & Mansell, T. J. (2019). Prebiotics: tools to manipulate the gut microbiome and metabolome. *Journal of industrial microbiology & biotechnology*, 1-15.

European Commision. (2010). Research area of Food, Agriculture & Fisheries & Biotechnology. Functional Foods. URL ftp://ftp.cordis.europa.eu/pub/fp7/kbbe/docs/functional-foods_en.pdf. Last accessed (01/10/2019)

Fabris, A., Biasioli, F., Granitto, P. M., Aprea, E., Cappellin, L., Schuhfried, E., Soukoulis, C., Märk, T. D., Gasperi, F., & Endrizzi, I. (2010). PTR-TOF-MS and data-mining methods for rapid characterization of agro-industrial samples:

influence of milk storage conditions on the volatile compounds profile of Trentingrana cheese. *Journal of mass spectrometry*, *45*(9), 1065-1074.

Falony, G., Lazidou, K., Verschaeren, A., Weckx, S., Maes, D., & De Vuyst, L. (2009). In vitro kinetic analysis of fermentation of prebiotic inulin-type fructans by Bifidobacterium species reveals four different phenotypes. *Appl. Environ. Microbiol.*, *75*(2), 454-461.

Fanaro, S., Jelinek, J., Stahl, B., Boehm, G., Kock, R., & Vigi, V. (2005). Acidic oligosaccharides from pectin hydrolysate as new component for infant formulae: effect on intestinal flora, stool characteristics, and pH. *Journal of pediatric gastroenterology and nutrition*, *41*(2), 186-190.

Femenia, A., Robertson, J. A., Waldron, K. W., & Selvendran, R. R. (1998). Cauliflower (Brassica oleracea L), globe artichoke (Cynara scolymus) and chicory witloof (Cichorium intybus) processing by-products as sources of dietary fibre. *Journal of the Science of Food and Agriculture*, *77*(4), 511-518.

Fernandes, S. S., Coelho, M. S., & de las Mercedes Salas-Mellado, M. (2019). Bioactive Compounds as Ingredients of Functional Foods: Polyphenols, Carotenoids, Peptides From Animal and Plant Sources New. In *Bioactive Compounds* (pp. 129-142). Woodhead Publishing.

Ferracane, R., Pellegrini, N., Visconti, A., Graziani, G., Chiavaro, E., Miglio, C., & Fogliano, V. (2008). Effects of different cooking methods on antioxidant profile, antioxidant capacity, and physical characteristics of artichoke. *Journal of agricultural and food chemistry*, *56*(18), 8601-8608.

Ferreira-Lazarte, A., Moreno, F. J., Cueva, C., Gil-Sánchez, I., & Villamiel, M. (2019). Behaviour of citrus pectin during its gastrointestinal digestion and fermentation in a dynamic simulator (simgi®). *Carbohydrate polymers*, *207*, 382-390.

Ferreira-Lazarte, A., Kachrimanidou, V., Villamiel, M., Rastall, R. A., & Moreno, F. J. (2018). In vitro fermentation properties of pectins and enzymatic-modified pectins obtained from different renewable bioresources. *Carbohydrate polymers*, *199*, 482-491.

Fishman, M. L., Chau, H. K., Hotchkiss Jr, A. T., White, A., Garcia, R. A., & Cooke, P. H. (2019). Effect of long term cold storage and microwave extraction time on the physical and chemical properties of citrus pectin. *Food hydrocolloids*, *92*, 104-116.

Fishman, M. L., El-Atawy, Y. S., Sondey, S. M., Gillespie, D. T., & Hicks, K. B. (1991). Component and global average radii of gyration of pectins from various sources. *Carbohydrate polymers*, *15*(1), 89-104.

Fissore, E. N., Santo Domingo, C., Pujol, C. A., Damonte, E. B., Rojas, A. M., & Gerschenson, L. N. (2014). Upgrading of residues of bracts, stems and hearts of Cynara cardunculus L. var. scolymus to functional fractions enriched in soluble fiber. *Food & function*, *5*(3), 463-470.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1), 1-22. URL http://www.jstatsoft.org/v33/i01/. Last accessed: 24/06/2019

Fu, D. D., Wang, Q. H., Ma, M. H., Ma, Y. X., & Wang, B. (2019). Nondestructive prediction modeling of S-ovalbumin content in stored eggs based on hyperspectral fusion information. *Journal of Food Process Engineering*, *42*(3), e13015.

Funami, T., Nakauma, M., Ishihara, S., Tanaka, R., Inoue, T., & Phillips, G. O. (2011). Structural modifications of sugar beet pectin and the relationship of structure to functionality. *Food Hydrocolloids*, *25*(2), 221-229.

Füzfai, Z., Kovács, E., & Molnár-Perl, I. (2004). Identification and quantitation of the main constituents of sour cherries: Simultaneously, as their trimethylsilyl derivatives, by gas chromatography-mass spectrometry. *Chromatographia*, *60*(1), S143-S151.

Gallardo-Escamilla, F. J., Kelly, A. L., & Delahunty, C. M. (2007). Mouthfeel and flavour of fermented whey with added hydrocolloids. *International Dairy Journal*, *17*(4), 308-315.

Gama, R., Van Dyk, J. S., & Pletschke, B. I. (2015). Optimization of enzymatic hydrolysis of apple pomace for production of biofuel and biorefinery chemicals using commercial enzymes. *3 Biotech*, *5*(6), 1075-1087.

Gamboa-Santos, J., Montilla, A., Soria, A. C., & Villamiel, M. (2012). Effects of conventional and ultrasound blanching on enzyme inactivation and carbohydrate content of carrots. *European Food Research and Technology*, *234*(6), 1071-1079.

Gan, C. Y., & Latiff, A. A. (2011). Extraction of antioxidant pectic-polysaccharide from mangosteen (Garcinia mangostana) rind: Optimization using response surface methodology. *Carbohydrate Polymers*, *83*(2), 600-607.

Ganje, M., Jafari, S. M., Tamadon, A. M., Niakosari, M., & Maghsoudlou, Y. (2019). Mathematical and fuzzy modeling of limonene release from amylose nanostructures and evaluation of its release kinetics. *Food Hydrocolloids*, *95*, 186-194.

Garna, H., Mabon, N., Nott, K., Wathelet, B., & Paquot, M. (2006). Kinetic of the hydrolysis of pectin galacturonic acid chains and quantification by ionic chromatography. *Food Chemistry*, *96*(3), 477-484.

Garna, H., Mabon, N., Wathelet, B., & Paquot, M. (2004). New method for a two-step hydrolysis and chromatographic analysis of pectin neutral sugar chains. *Journal of agricultural and food chemistry*, *52*(15), 4652-4659.

Georgiev, Y. N., Paulsen, B. S., Kiyohara, H., Ciz, M., Ognyanov, M. H., Vasicek, O., Rise, F., Denev, P. N., Lojek, A., Batsalova, T. G., Dzhambazov, B. M., Yamada, H., Lund., R., Barsett, H., Krastanov, A. I., Yanakieva, I. Z., & Kratchanova, M.G. (2017a). Tilia tomentosa pectins exhibit dual mode of action on phagocytes as β-glucuronic acid monomers are abundant in their rhamnogalacturonans I. *Carbohydrate polymers*, *175*, 178-191.

Georgiev, Y. N., Paulsen, B. S., Kiyohara, H., Ciz, M., Ognyanov, M. H., Vasicek, O., Rise, F., Denev, P. N., Yamada, H., Lojek, A., Kussovski, V., Barsett, H., Krastanov, A. I., Yanakieva, I. Z., & Kratchanova, M.G. (2017b). The common lavender (Lavandula angustifolia Mill.) pectic polysaccharides modulate phagocytic leukocytes and intestinal Peyer's patch cells. *Carbohydrate polymers*, *174*, 948-959.

Gerbst, A. G., Grachev, A. A., Ustuzhanina, N. E., Nifantiev, N. E., Vyboichtchik, A. A., Shashkov, A. S., & Usov, A. I. (2010). Application of artificial neural networks for analysis of 13C NMR spectra of fucoidans. *Journal of Carbohydrate Chemistry*, *29*(2), 92-102.

Gerschenson, L. N. (2017). The production of galacturonic acid enriched fractions and their functionality. *Food hydrocolloids*, *68*, 23-30.

Gertheiss, J., & Tutz, G. (2009). Supervized feature selection in mass spectrometry-based proteomic profiling by blockwise boosting. *Bioinformatics*, *25*(8), 1076-1077.

Geurts, P., Fillet, M., De Seny, D., Meuwis, M. A., Malaise, M., Merville, M. P., & Wehenkel, L. (2005). Proteomic mass spectra classification using decision tree based ensemble methods. *Bioinformatics*, *21*(14), 3138-3145.

Gibson, G. R., & Roberfroid, M. B. (1995). Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. *The Journal of nutrition*, *125*(6), 1401-1412.

Golden, C. E., Rothrock Jr, M. J., & Mishra, A. (2019). Comparison between random forest and gradient boosting machine methods for predicting Listeria spp. prevalence in the environment of pastured poultry farms. *Food Research International*, *122*, 47-55.

Gómez, B., Gullón, B., Yáñez, R., Schols, H., & Alonso, J. L. (2016). Prebiotic potential of pectins and pectic oligosaccharides derived from lemon peel wastes and sugar beet pulp: A comparative evaluation. *Journal of Functional Foods*, *20*, 108-121.

Gómez, B., Gullón, B., Remoroza, C., Schols, H. A., Parajó, J. C., & Alonso, J. L. (2014). Purification, characterization, and prebiotic properties of pectic oligosaccharides from orange peel wastes. *Journal of agricultural and food chemistry*, *62*(40), 9769-9782.

Gómez, B., Gullón, B., Yáñez, R., Parajó, J. C., & Alonso, J. L. (2013). Pectic oligosacharides from lemon peel wastes: production, purification, and chemical characterization. *Journal of agricultural and food chemistry*, *61*(42), 10043-10053.

Gosav, S., Praisler, M., & Birsa, M. L. (2011). Principal component analysis coupled with artificial neural networks—A combined technique classifying small

molecular structures using a concatenated spectral database. *International journal of molecular sciences*, *12*(10), 6668-6684.

Gostin, A. I., & Waisundara, V. Y. (2019). Edible flowers as functional food: A review on artichoke (Cynara cardunculus L.). *Trends in Food Science & Technology*.

Grassino, A. N., Barba, F. J., Brnčić, M., Lorenzo, J. M., Lucini, L., & Brnčić, S. R. (2018). Analytical tools used for the identification and quantification of pectin extracted from plant food matrices, wastes and by-products: A review. *Food chemistry*, *266*, 47-55.

Gu, S., Wang, J., & Wang, Y. (2019). Early discrimination and growth tracking of Aspergillus spp. contamination in rice kernels using electronic nose. *Food chemistry*, *292*, 325-335.

Gullón, B., Gómez, B., Martínez-Sabajanes, M., Yáñez, R., Parajó, J. C., & Alonso, J. L. (2013). Pectic oligosaccharides: Manufacture and functional properties. *Trends in Food Science & Technology*, *30*(2), 153-161.

Gummadi, S. N., & Panda, T. (2003). Purification and biochemical properties of microbial pectinases—a review. *Process biochemistry*, *38*(7), 987-996.

Gur, J., Mawuntu, M., & Martirosyan, D. (2018). FFC's Advancement of Functional Food Definition. *Functional Foods in Health and Disease*, *8*(7), 385-297.

Herbstreith & Fox. (2018). The Specialists for Pectin. URL http://www.herbstreith-fox.de/index.php?id=54&L=1. Last accessed: 30/07/18

Holck, J., Hotchkiss Jr, A. T., Meyer, A. S., Mikkelsen, J. D., & Rastall, R. A. (2014). Production and bioactivity of pectic oligosaccharides from fruit and vegetable biomass. *Food Oligosaccharides: Production, Analysis and Bioactivity. Oxford: John Wiley & Sons, Ltd*, 76-87.

Holck, J., Lorentzen, A., Vigsnæs, L. K., Licht, T. R., Mikkelsen, J. D., & Meyer, A. S. (2011). Feruloylated and nonferuloylated arabino-oligosaccharides from sugar beet pectin selectively stimulate the growth of Bifidobacterium spp. in human fecal in vitro fermentations. *Journal of agricultural and food chemistry*, *59*(12), 6511-6519.

Hu, Q., Lü, M. H., & Deng, M. M. (2018). 125-Effects of Fecal Microbiota Transplantation on DSS-Induced Colitis Mice May Partly Owing to Enhanced Polarization of Macrophage M2. *Gastroenterology*, *154*(6), S-33.

Huang, B., Zhao, K., Zhang, Z., Liu, F., Hu, H., & Pan, S. (2018). Changes on the rheological properties of pectin-enriched mango nectar by high intensity ultrasound. *LWT*, *91*, 414-422.

Im Cho, Y., & Palvanov, A. (2019). A New Machine Learning Algorithm for Weather Visibility and Food Recognition. *Journal of Robotics, Networking and Artificial Life*, *6*(1), 12-17.

Ishii, T., Ichita, J., Matsue, H., Ono, H., & Maeda, I. (2002). Fluorescent labeling of pectic oligosaccharides with 2-aminobenzamide and enzyme assay for pectin. *Carbohydrate research*, *337*(11), 1023-1032.

Ishisono, K., Yabe, T., & Kitaguchi, K. (2017). Citrus pectin attenuates endotoxin shock via suppression of Toll-like receptor signaling in Peyer's patch myeloid cells. *The Journal of nutritional biochemistry*, *50*, 38-45.

Jagtap, S., Bhatt, C., Thik, J., & Rahimifard, S. (2019). Monitoring Potato Waste in Food Manufacturing Using Image Processing and Internet of Things Approach. *Sustainability*, *11*(11), 3173.

Jeddou, K. B., Bouaziz, F., Helbert, C. B., Nouri-Ellouz, O., Maktouf, S., Ellouz-Chaabouni, S., & Ellouz-Ghorbel, R. (2018). Structural, functional, and

biological properties of potato peel oligosaccharides. *International journal of biological macromolecules*, *112*, 1146-1155.

Jeurink, P. V., van Esch, B. C., Rijnierse, A., Garssen, J., & Knippels, L. M. (2013). Mechanisms underlying immune effects of dietary oligosaccharides. *The American journal of clinical nutrition*, *98*(2), 572S-577S.

Jha, S. K., Josheski, F., Marina, N., & Hayashi, K. (2016). GC–MS characterization of body odour for identification using artificial neural network classifiers fusion. *International Journal of Mass Spectrometry*, *406*, 35-47.

Jiang, T., Gao, X., Wu, C., Tian, F., Lei, Q., Bi, J., Xie, B., Wang, H. Y., Chen, S., & Wang, X. (2016). Apple-derived pectin modulates gut microbiota, improves gut barrier function, and attenuates metabolic endotoxemia in rats with diet-induced obesity. *Nutrients*, *8*(3), 126.

Jiménez-Carvelo, A. M., González-Casado, A., Bagur-González, M. G., & Cuadros-Rodríguez, L. (2019). Alternative data mining/machine learning methods for the analytical evaluation of food quality and authenticity–A review. *Food research international*.

Jin, M., Wang, Y., Yang, X., Yin, H., Nie, S., & Wu, X. (2019). Structure characterization of a polysaccharide extracted from noni (Morinda citrifolia L.) and its protective effect against DSS-induced bowel disease in mice. *Food hydrocolloids*, *90*, 189-197.

Johnson, I.; Hahsler, M. (2018). arulesCBA: Classification Based on Association Rules. *R package version 1.1.3-1*. URL https://CRAN.R-project.org/package=arulesCBA. Last accessed: 16/12/2018

Jouini, M., Abdelhamid, A., Chaouch, M. A., le Cerf, D., Bouraoui, A., Majdoub, H., & Jannet, H. B. (2018). Physico-chemical characterization and pharmacological

activities of polysaccharides from Opuntia microdasys var. rufida cladodes. *International journal of biological macromolecules*, *107*, 1330-1338.

Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., & MacCoss, M. J. (2007). Semi-supervized learning for peptide identification from shotgun proteomics datasets. *Nature methods*, *4*(11), 923.

Karaca, O. B. (2013). Effects of different prebiotic stabilisers and types of molasses on physicochemical, sensory, colour and mineral characteristics of probiotic set yoghurt. *International Journal of Dairy Technology*, *66*(4), 490-497.

Karboune, S., & Khodaei, N. (2016). Structures, isolation and health-promoting properties of pectic polysaccharides from cell wall-rich food by-products: a source of functional ingredients. *Current Opinion in Food Science*, *8*, 50-55.

Khangwal, I., & Shukla, P. (2019). Prospecting prebiotics, innovative evaluation methods, and their health applications: a review. *3 Biotech*, *9*(5), 187.

Kim, M., Friesen, L., Park, J., Kim, H. M., & Kim, C. H. (2018). Microbial metabolites, short-chain fatty acids, restrain tissue bacterial load, chronic inflammation, and associated cancer in the colon of mice. *European journal of immunology*, *48*(7), 1235-1247.

Kollia, E., Markaki, P., Zoumpoulakis, P., & Proestos, C. (2017). Antioxidant activity of Cynara scolymus L. and Cynara cardunculus L. extracts obtained by different extraction techniques. *Natural product research*, *31*(10), 1163-1167.

Konwarh, R., Pramanik, S., Kalita, D., Mahanta, C. L., & Karak, N. (2012). Ultrasonication–A complementary ‘green chemistry’tool to biocatalysis: A laboratory-scale study of lycopene extraction. *Ultrasonics Sonochemistry*, *19*(2), 292-299.

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervized machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, *160*, 3-24.

Krzeminski, A., Prell, K. A., Busch-Stockfisch, M., Weiss, J., & Hinrichs, J. (2014). Whey protein–pectin complexes as new texturising elements in fat-reduced yoghurt systems. *International Dairy Journal*, *36*(2), 118-127.

Kuivanen, J., Biz, A., & Richard, P. (2019). Microbial hexuronate catabolism in biotechnology. *AMB Express*, *9*(1), 16.

Larsen, N., Bussolo de Souza, C., Krych, L., Barbosa Cahu, T., Wiese, M., Kot, W., Meyer Hansen, K., Blennow, A., Venema, K., & Jespersen, L. (2019). Potential of pectins to beneficially modulate the gut microbiota depends on their structural properties. *Frontiers in microbiology*, *10*, 223.

Lattanzio, V., Kroon, P. A., Linsalata, V., & Cardinali, A. (2009). Globe artichoke: a functional food and source of nutraceutical ingredients. *Journal of functional foods*, *1*(2), 131-144.

Lattanzio, V., & van Sumere, C. F. (1987). Changes in phenolic compounds during the development and cold storage of artichoke (Cynara scolymus L.) heads. *Food Chemistry*, *24*(1), 37-50.

LeDell, E., Gill, N., Aiello, S., Fu, A., Candel, A., Click, C., Kraljevic, T., Nykodym, T., Aboyoun, P., Kurka, M., & and Malohlava, M. (2019). h2o: R Interface for 'H2O'. *R package version 3.22.1.1*. URL https://CRAN.R-project.org/package=h2o. Last accessed: 26/08/2019

Lee, J. H., Lee, Y. K., Choi, Y. R., Park, J., Jung, S. K., & Chang, Y. H. (2018). The characterization, selenylation and anti-inflammatory activity of pectic

polysaccharides extracted from Ulmus pumila L. *International journal of biological macromolecules*, *111*, 311-318.

Leijdekkers, A. G., Huang, J. H., Bakx, E. J., Gruppen, H., & Schols, H. A. (2015). Identification of novel isomeric pectic oligosaccharides using hydrophilic interaction chromatography coupled to traveling-wave ion mobility mass spectrometry. *Carbohydrate research*, *404*, 1-8.

Leivas, C. L., Nascimento, L. F., Barros, W. M., Santos, A. R., Iacomini, M., & Cordeiro, L. M. (2016). Substituted galacturonan from starfruit: Chemical structure and antinociceptive and anti-inflammatory effects. *International journal of biological macromolecules*, *84*, 295-300.

Leroux, J., Langendorff, V., Schick, G., Vaishnav, V., & Mazoyer, J. (2003). Emulsion stabilizing properties of pectin. *Food Hydrocolloids*, *17*(4), 455-462.

Li, J., Li, S., Liu, S., Wei, C., Yan, L., Ding, T., Linhardt, J., Liu, D., Ye, X., & Chen, S. (2019a). Pectic oligosaccharides hydrolyzed from citrus canning processing water by Fenton reaction and their antiproliferation potentials. *International journal of biological macromolecules*, *124*, 1025-1032.

Li, T., Chen, X., Huang, Z., Xie, W., Tong, C., Bao, R., Sun, X., Li, W., & Li, S. (2019b). Pectin oligosaccharide from hawthorn fruit ameliorates hepatic inflammation via NF-κB inactivation in high-fat diet fed mice. *Journal of Functional Foods*, *57*, 345-350.

Li, P. J., Xia, J. L., Nie, Z. Y., & Shan, Y. (2016). Pectic oligosaccharides hydrolyzed from orange peel by fungal multi-enzyme complexes and their prebiotic and antibacterial potentials. *LWT-Food Science and Technology*, *69*, 203-210.

Li, X., He, X., Lv, Y., & He, Q. (2014). Extraction and functional properties of water-soluble dietary fiber from apple pomace. *Journal of food process engineering*, *37*(3), 293-298.

Li, Y., Liu, L., Niu, Y., Feng, J., Sun, Y., Kong, X., Chen, Y., Chen, X., Gan, H., Cao, S., & Mei, Q. (2012). Modified apple polysaccharide prevents against tumorigenesis in a mouse model of colitis-associated colon cancer: role of galectin-3 and apoptosis in cancer prevention. *European journal of nutrition*, *51*(1), 107-117.

Li, X., Li, J., & Yao, X. (2007). A wavelet-based data pre-processing analysis approach in mass spectrometry. *Computers in Biology and Medicine*, *37*(4), 509-516.

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*, 18-22.

Liew, S. Q., Chin, N. L., Yusof, Y. A., & Sowndhararajan, K. (2016). Comparison of acidic and enzymatic pectin extraction from passion fruit peels and its gel properties. *Journal of Food Process Engineering*, *39*(5), 501-511.

Lim, D. K., Long, N. P., Mo, C., Dong, Z., Cui, L., Kim, G., & Kwon, S. W. (2017). Combination of mass spectrometry-based targeted lipidomics and supervized machine learning algorithms in detecting adulterated admixtures of white rice. *Food research international*, *100*, 814-821.

Lin, Z., Gonçalves, C. M. V., Dai, L., Lu, H. M., Huang, J. H., Ji, H., Wang, D., Yi, L., & Liang, Y. Z. (2014). Exploring metabolic syndrome serum profiling based on gas chromatography mass spectrometry and random forest models. *Analytica chimica acta*, *827*, 22-27.

Liu, R., Zhang, G., Sun, M., Pan, X., & Yang, Z. (2019a). Integrating a generalized data analysis workflow with the Single-probe mass spectrometry experiment for single cell metabolomics. *Analytica chimica acta*, *1064*, 71-79.

Liu, W., Zhao, P., Wu, C., Liu, C., Yang, J., & Zheng, L. (2019b). Rapid determination of aflatoxin B1 concentration in soybean oil using terahertz spectroscopy with chemometric methods. *Food chemistry*, *293*, 213-219.

Liu, J., Liu, L., Guo, W., Fu, M., Yang, M., Huang, S., Zhang, & Liu, Y. (2019c). A new methodology for sensory quality assessment of garlic based on metabolomics and an artificial neural network. *RSC Advances*, *9*(31), 17754-17765.

Liu, Z., Pi, F., Guo, X., Guo, X., & Yu, S. (2019d). Characterization of the structural and emulsifying properties of sugar beet pectins obtained by sequential extraction. *Food hydrocolloids*, *88*, 31-42.

Liu, Y., Gong, G., Zhang, J., Jia, S., Li, F., Wang, Y., & Wu, S. (2014). Response surface optimization of ultrasound-assisted enzymatic extraction polysaccharides from Lycium barbarum. *Carbohydrate polymers*, *110*, 278-284.

Liu, L., Li, Y. H., Niu, Y. B., Sun, Y., Guo, Z. J., Li, Q., Li, C., Li, J., S., Feng, Cao, S., & Mei, Q. B. (2010). An apple oligogalactan prevents against inflammation and carcinogenesis by targeting LPS/TLR4/NF-κB pathway in a mouse model of colitis-associated colon cancer. *Carcinogenesis*, *31*(10), 1822-1832.

Llorach, R., Espin, J. C., Tomas-Barberan, F. A., & Ferreres, F. (2002). Artichoke (Cynara scolymus L.) byproducts as a potential source of health-promoting antioxidant phenolics. *Journal of Agricultural and Food Chemistry*, *50*(12), 3458-3464.

Longnecker, K., & Kujawinski, E. B. (2017). Mining mass spectrometry data: Using new computational tools to find novel organic compounds in complex environmental mixtures. *Organic geochemistry*, *110*, 92-99.

López-Molina, D., Navarro-Martínez, M. D., Rojas-Melgarejo, F., Hiner, A. N., Chazarra, S., & Rodríguez-López, J. N. (2005). Molecular properties and prebiotic effect of inulin obtained from artichoke (Cynara scolymus L.). *Phytochemistry*, *66*(12), 1476-1484.

López-Ruiz, R., Romero-González, R., Ortega-Carrasco, E., & Garrido Frenich, A. (2019). Dissipation studies of famoxadone in vegetables under greenhouse conditions using liquid chromatography coupled to high-resolution mass spectrometry: putative elucidation of a new metabolite. *Journal of the Science of Food and Agriculture*, *99*, 5368-5376.

Ma, X., Wang, D., Chen, W., Ismail, B. B., Wang, W., Lv, R., Ding, T.; Ye, X.; & Liu, D. (2018). Effects of ultrasound pretreatment on the enzymolysis of pectin: Kinetic study, structural characteristics and anti-cancer activity of the hydrolysates. *Food hydrocolloids*, *79*, 90-99.

Machado, M. T., Eça, K. S., Vieira, G. S., Menegalli, F. C., Martínez, J., & Hubinger, M. D. (2015). Prebiotic oligosaccharides from artichoke industrial waste: evaluation of different extraction methods. *Industrial Crops and Products*, *76*, 141-148.

Magne, F., Hachelaf, W., Suau, A., Boudraa, G., Bouziane-Nedjadi, K., Rigottier-Gois, L., Touhami, M., Desjeux, J. F., & Pochart, P. (2008). Effects on faecal microbiota of dietary and acidic oligosaccharides in children during partial formula feeding. *Journal of pediatric gastroenterology and nutrition*, *46*(5), 580-588.

Majka, M. (2019). naivebayes: High Performance Implementation of the Naive Bayes Algorithm. *R package version 0.9.5*. URL https://CRAN.R-project.org/package=naivebayes. Last accessed: 24/06/2019

Mandalari, G., Palop, C. N., Tuohy, K., Gibson, G. R., Bennett, R. N., Waldron, K. W., & Faulds, C. B. (2007). In vitro evaluation of the prebiotic activity of a pectic oligosaccharide-rich extract enzymatically derived from bergamot peel. *Applied Microbiology and Biotechnology*, *73*(5), 1173-1179.

Marenda, F. R. B., Mattioda, F., Demiate, I. M., de Francisco, A., de Oliveira Petkowicz, C. L., Canteri, M. H. G., & Amboni, R. D. D. M. C. (2019). Advances in Studies Using Vegetable Wastes to Obtain Pectic Substances: A Review. *Journal of polymers and the environment*, *27*(3), 549-560.

Marić, M., Grassino, A. N., Zhu, Z., Barba, F. J., Brnčić, M., & Brnčić, S. R. (2018). An overview of the traditional and innovative approaches for pectin extraction from plant food wastes and by-products: Ultrasound-, microwaves-, and enzyme-assisted extraction. *Trends in food science & technology*, *76*, 28-37.

Martínez Sabajanes, M., Yáñez, R., Alonso, J. L., & Parajó, J. C. (2012). Pectic oligosaccharides production from orange peel waste by enzymatic hydrolysis. *International journal of food science & technology*, *47*(4), 747-754.

Martinez, M., Gullon, B., Yanez, R., Alonso, J. L., & Parajó, J. C. (2009a). Direct enzymatic production of oligosaccharide mixtures from sugar beet pulp: experimental evaluation and mathematical modeling. *Journal of agricultural and food chemistry*, *57*(12), 5510-5517.

Martínez, M., Gullón, B., Schols, H. A., Alonso, J. L., & Parajó, J. C. (2009b). Assessment of the production of oligomeric compounds from sugar beet pulp. *Industrial & Engineering Chemistry Research*, *48*(10), 4681-4687.

Martínez-Villaluenga, C., Cardelle-Cobas, A., Corzo, N., & Olano, A. (2008). Study of galactooligosaccharide composition in commercial fermented milks. *Journal of food composition and analysis*, *21*(7), 540-544.

Maxwell, E. G., Colquhoun, I. J., Chau, H. K., Hotchkiss, A. T., Waldron, K. W., Morris, V. J., & Belshaw, N. J. (2015). Rhamnogalacturonan I containing homogalacturonan inhibits colon cancer cell proliferation by decreasing ICAM1 expression. *Carbohydrate polymers*, *132*, 546-553.

Maxwell, E. G., Belshaw, N. J., Waldron, K. W., & Morris, V. J. (2012). Pectin–an emerging new bioactive food polysaccharide. *Trends in Food Science & Technology*, *24*(2), 64-73.

Meinshausen, N. (2017). quantregForest: Quantile Regression Forests. *R package version 1.3-7*. https://CRAN.R-project.org/package=quantregForest. Last accessed: 02/07/2019

Meringer, M., & Schymanski, E. (2013). Small molecule identification with MOLGEN and mass spectrometry. *Metabolites*, *3*(2), 440-462.

Miller, G. L. (1959). Modified DNS method for reducing sugars. *Analytical chemistry*, *31*(3), 426-428.

Min, B., Lim, J., Ko, S., Lee, K. G., Lee, S. H., & Lee, S. (2011). Environmentally friendly preparation of pectins from agricultural byproducts and their structural/rheological characterization. *Bioresource Technology*, *102*(4), 3855-3860.

Mohnen, D. (2008). Pectin structure and biosynthesis. *Current opinion in plant biology*, *11*(3), 266-277.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*. MIT press.

Moncayo, S., Manzoor, S., Rosales, J. D., Anzano, J., & Caceres, J. O. (2017). Qualitative and quantitative analysis of milk for the detection of adulteration by Laser Induced Breakdown Spectroscopy (LIBS). *Food chemistry*, *232*, 322-328.

Moreno, F. J., Corzo, N., Montilla, A., Villamiel, M., & Olano, A. (2017). Current state and latest advances in the concept, production and functionality of prebiotic oligosaccharides. *Current Opinion in Food Science*, *13*, 50-55.

Morris, V. J., Belshaw, N. J., Waldron, K. W., & Maxwell, E. G. (2013). The bioactivity of modified pectin fragments. *Bioactive Carbohydrates and Dietary Fibre*, *1*(1), 21-37.

Muñoz-Almagro, N., Rico-Rodriguez, F., Wilde, P. J., Montilla, A., & Villamiel, M. (2018). Structural and technological characterization of pectin extracted with sodium citrate and nitric acid from sunflower heads. *Electrophoresis*, *39*(15), 1984-1992.

Muñoz-Almagro, N., Montilla, A., Moreno, F. J., & Villamiel, M. (2017). Modification of citrus and apple pectin by power ultrasound: Effects of acid and enzymatic treatment. *Ultrasonics sonochemistry*, *38*, 807-819.

Mussatto, S. I., & Mancilha, I. M. (2007). Non-digestible oligosaccharides: a review. *Carbohydrate polymers*, *68*(3), 587-597.

Myers, R. H., Montgomery, D. C., Vining, G. G., Borror, C. M., & Kowalski, S. M. (2004). Response surface methodology: a retrospective and literature survey. *Journal of quality technology*, *36*(1), 53-77.

Mzoughi, Z., Abdelhamid, A., Rihouey, C., Le Cerf, D., Bouraoui, A., & Majdoub, H. (2018). Optimized extraction of pectin-like polysaccharide from Suaeda fruticosa leaves: Characterization, antioxidant, anti-inflammatory and analgesic activities. *Carbohydrate polymers*, *185*, 127-137.

Nadar, S. S., Rao, P., & Rathod, V. K. (2018). Enzyme assisted extraction of biomolecules as an approach to novel extraction technology: A review. *Food Research International*, *108*, 309-330.

Nguyen, D. H., Nguyen, C. H., & Mamitsuka, H. (2018). Recent advances and prospects of computational methods for metabolite identification: a review with emphasis on machine learning approaches. *Briefings in bioinformatics*, *0*(0), 1-16.

Nguyen, T. T. T., & Le, V. V. M. (2013). Effects of ultrasound on cellulolytic activity of cellulase complex. *International Food Research Journal*, *20*(2), 557-563.

Ognyanov, M., Remoroza, C., Schols, H. A., Georgiev, Y., Kratchanova, M., & Kratchanov, C. (2016). Isolation and structure elucidation of pectic polysaccharide from rose hip fruits (Rosa canina L.). *Carbohydrate polymers*, *151*, 803-811.

Olano-Martin, E., Gibson, G. R., & Rastall, R. A. (2002). Comparison of the in vitro bifidogenic properties of pectins and pectic-oligosaccharides. *Journal of Applied Microbiology*, *93*(3), 505-511.

Olano-Martin, E., Mountzouris, K. C., Gibson, G. R., & Rastall, R. A. (2001). Continuous production of pectic oligosaccharides in an enzyme membrane reactor. *Journal of Food Science*, *66*(7), 966-971.

Olivera, P., Danese, S., Jay, N., Natoli, G., & Peyrin-Biroulet, L. (2019). Big data in IBD: a look into the future. *Nature Reviews Gastroenterology & Hepatology*, *16*, 312–321.

Ordaz-Ortiz, J. J., Devaux, M. F., & Saulnier, L. (2005). Classification of wheat varieties based on structural features of arabinoxylans as revealed by

endoxylanase treatment of flour and grain. *Journal of Agricultural and Food Chemistry*, *53*(21), 8349-8356.

Pabst, M., Fischl, R. M., Brecker, L., Morelle, W., Fauland, A., Köfeler, H., Altmann, F., & Léonard, R. (2013). Rhamnogalacturonan II structure shows variation in the side chains monosaccharide composition and methylation status within and across different plant species. *The Plant Journal*, *76*(1), 61-72.

Pacheco, M. T., Vezza, T., Diez-Echave, P., Utrilla, P., Villamiel, M., & Moreno, F. J. (2018). Anti-inflammatory bowel effect of industrial orange by-products in DSS-treated mice. *Food & function*, *9*(9), 4888-4896.

Pagan, J., Ibarz, A., Llorca, M., Pagan, A., & Barbosa-Cánovas, G. V. (2001). Extraction and characterization of pectin from stored peach pomace. *Food Research International*, *34*(7), 605-612.

Panchev, I. N., Kirtchev, N. A., & Kratchanov, C. G. (1994). On the production of low esterified pectins by acid maceration of pectic raw materials with ultrasound treatment. *Food Hydrocolloids*, *8*(1), 9-17.

Pardo-Rueda, A. J., Quintero-Ramos, A., Genovese, D. B., Camacho-Dávila, A., Zepeda-Rodríguez, A., Contreras-Esquivel, J. C., & Bizarro, A. P. (2015). Efficient extraction of fructans from sotol plant (Dasylirion leiophyllum) enhanced by a combination of enzymatic and sonothermal treatments. *Food and Bioproducts Processing*, *94*, 398-404.

Park, J., Kim, M., Kang, S. G., Jannasch, A. H., Cooper, B., Patterson, J., & Kim, C. H. (2015). Short-chain fatty acids induce both effector and regulatory T cells by suppression of histone deacetylases and regulation of the mTOR–S6K pathway. *Mucosal immunology*, *8*(1), 80.

Pedersen, H. K., Gudmundsdottir, V., Nielsen, H. B., Hyotylainen, T., Nielsen, T., Jensen, B. A., Forslund, K., Hildebrand, F, Prifti, E., Falony, G., Le Chatelier, E., Levenez, F., Doré, J., Mattila, I., Plichta, R. d. Pöhö, P., Hellgren L. I., Arumugam, M., Sunagawa, S., Vieira-Silva, S., Jorgensen, T., Holm, J. B., Trost, K., Consortium, M., Kristiansen, K., Brix, S., Raes, J., Wang, J., Hansen, T., Bork, P., Brunak, S., Oresic, M., Ehrlich, S. D., & Pedersen, O. (2016). Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature*, *535*(7612), 376–381.

Peng, F., Peng, P., Xu, F., & Sun, R. C. (2012). Fractional purification and bioconversion of hemicelluloses. *Biotechnology advances*, *30*(4), 879-903.

Peters, A., & Hothorn, T. (2018). ipred: Improved Predictors. *R package version 0.9-8*. https://CRAN.R-project.org/package=ipred Last accessed: 02/07/2019

Peterson, G. (1974). Gas-chromatographic analysis of sugars and related hydroxy acids as acyclic oxime and ester trimethylsilyl derivatives. *Carbohydrate Research*, *33*(1), 47-61.

Pianta, A., Arvikar, S., Strle, K., Drouin, E. E., Wang, Q., Costello, C. E., & Steere, A. C. (2017). Evidence of the immune relevance of Prevotella copri, a gut microbe, in patients with rheumatoid arthritis. *Arthritis & rheumatology*, *69*(5), 964-975.

Playne, M. J., & Crittenden, R. (1996). Commercially available oligosaccharides. *International Dairy Federation*, Bulletin-FIL-IDF (Belgium).

Pontiggia, D., Ciarcianelli, J., Salvi, G., Cervone, F., De Lorenzo, G., & Mattei, B. (2015). Sensitive detection and measurement of oligogalacturonides in Arabidopsis. *Frontiers in plant science*, *6*, 258.

Poondla, V., Bandikari, R., Subramanyam, R., & Obulam, V. S. R. (2015). Low temperature active pectinases production by Saccharomyces cerevisiae isolate

and their characterization. *Biocatalysis and Agricultural Biotechnology*, *4*(1), 70-76.

Popov, S. V., & Ovodov, Y. S. (2013a). Polypotency of the immunomodulatory effect of pectins. *Biochemistry (Moscow)*, *78*(7), 823-835.

Popov, S. V., Markov, P. A., Popova, G. Y., Nikitina, I. R., Efimova, L., & Ovodov, Y. S. (2013b). Anti-inflammatory activity of low and high methoxylated citrus pectins. *Biomedicine & Preventive Nutrition*, *3*(1), 59-63.

Popov, S. V., Popova, G. Y., Paderin, N. M., Koval, O. A., Ovodova, R. G., & Ovodov, Y. S. (2007). Preventative antiinflammatory effect of potamogetonan, a pectin from the common pondweed Potamogeton natans L. *Phytotherapy Research: An International Journal Devoted to Pharmacological and Toxicological Evaluation of Natural Product Derivatives*, *21*(7), 609-614.

Prandi, B., Baldassarre, S., Babbar, N., Bancalari, E., Vandezande, P., Hermans, D., Bruggeman, G., Gatti, M., Elst, K., & Sforza, S. (2018). Pectin oligosaccharides from sugar beet pulp: molecular characterization and potential prebiotic activity. *Food & function*, *9*(3), 1557-1569.

Pu, J. B., Xia, B. H., Hu, Y. J., Zhang, H. J., Chen, J., Zhou, J., Liang, W. Q., & Xu, P. (2015). Multi-optimization of ultrasonic-assisted enzymatic extraction of Atratylodes macrocephala polysaccharides and antioxidants using response surface methodology and desirability function approach. *Molecules*, *20*(12), 22220-22235.

Rajagopal, H. M., Manjegowda, S. B., Serkad, C., & Dharmesh, S. M. (2018). A modified pectic polysaccharide from turmeric (Curcuma longa) with antiulcer effects via anti–secretary, mucoprotective and IL–10 mediated anti–

inflammatory mechanisms. *International journal of biological macromolecules*, *118*, 864-880.

Ralet, M. C., Crépeau, M. J., & Bonnin, E. (2008). Evidence for a blockwise distribution of acetyl groups onto homogalacturonans from a commercial sugar beet (Beta vulgaris) pectin. *Phytochemistry*, *69*(9), 1903-1909.

Ralet, M. C., Cabrera, J. C., Bonnin, E., Quéméner, B., Hellìn, P., & Thibault, J. F. (2005). Mapping sugar beet pectin acetylation pattern. *Phytochemistry*, *66*(15), 1832-1843.

Ramasamy, U. (2014). *The role of soluble and insoluble fibers during fermentation of Chicory root pulp*. Wageningen University.

Rehman, H. U., Siddique, N. N., Aman, A., Nawaz, M. A., Baloch, A. H., & Qader, S. A. U. (2015). Morphological and molecular based identification of pectinase producing Bacillus licheniformis from rotten vegetable. *Journal of Genetic Engineering and Biotechnology*, *13*(2), 139-144.

Rico-Rodríguez, F., Serrato, J. C., Montilla, A., & Villamiel, M. (2018). Impact of ultrasound on galactooligosaccharides and gluconic acid production throughout a multienzymatic system. *Ultrasonics sonochemistry*, *44*, 177-183.

Roberfroid, M., Gibson, G. R., Hoyles, L., McCartney, A. L., Rastall, R., Rowland, I., Wolvers, D., Watzl, B., Szajewska, H., Stahl, B., Guarner, F., Respondek, F., Whelan, K., Coxam, V., Davicco, M. J., Léotoing, L., Wittrant, Y., Delzenne, N. M., Cani, P. D., Neyrinck, A. M., & Meheust, A. (2010). Prebiotic effects: metabolic and health benefits. *British Journal of Nutrition*, *104*(S2), S1-S63.

Roberfroid, M., & Slavin, J. (2000). Nondigestible oligosaccharides. *Critical Reviews in Food Science and Nutrition.*, *40*(6), 461-480.

Robert, C., Emaga, T. H., Wathelet, B., & Paquot, M. (2008). Effect of variety and harvest date on pectin extracted from chicory roots (Cichorium intybus L.). *Food chemistry*, *108*(3), 1008-1018.

Rojas, C., Duchowicz, P. R., & Castro, E. A. (2019). Foodinformatics: Quantitative Structure-Property Relationship Modeling of Volatile Organic Compounds in Peppers. *Journal of food science*, *84*(4), 770-781.

Rojas-Cherto, M., Peironcely, J. E., Kasper, P. T., van der Hooft, J. J., de Vos, R. C., Vreeken, R., Hankemeier, T., & Reijmers, T. (2012). Metabolite identification using automated comparison of high-resolution multistage mass spectral trees. *Analytical chemistry*, *84*(13), 5524-5534.

Roselló-Soto, E., Parniakov, O., Deng, Q., Patras, A., Koubaa, M., Grimi, N., Bousseta, N., Tiwari, B. K., Vorobiev, E., Lebovka, N., & Barba, F. J. (2016). Application of non-conventional extraction methods: Toward a sustainable and green production of valuable compounds from mushrooms. *Food Engineering Reviews*, *8*(2), 214-234.

Rossel, S., & Martínez Arbizu, P. (2018). Automatic specimen identification of Harpacticoids (Crustacea: Copepoda) using Random Forest and MALDI-TOF mass spectra, including a post hoc test for false positive discovery. *Methods in Ecology and Evolution*, *9*(6), 1421-1434.

Rubio-Arraez, S., Capella, J. V., Castelló, M. L., & Ortolá, M. D. (2016). Physicochemical characteristics of citrus jelly with non cariogenic and functional sweeteners. *Journal of food science and technology*, *53*(10), 3642-3650.

Ruiz-Aceituno, L., García-Sarrió, M. J., Alonso-Rodriguez, B., Ramos, L., & Sanz, M. L. (2016). Extraction of bioactive carbohydrates from artichoke (Cynara

scolymus L.) external bracts using microwave assisted extraction and pressurized liquid extraction. *Food chemistry*, *196*, 1156-1162.

Ruttkies, C., Neumann, S., & Posch, S. (2019a). Improving MetFrag with statistical learning of fragment annotations. *BMC bioinformatics*, *20*(1), 376.

Ruttkies, C., Schymanski, E. L., Strehmel, N., Hollender, J., Neumann, S., Williams, A. J., & Krauss, M. (2019b). Supporting non-target identification by adding hydrogen deuterium exchange MS/MS capabilities to MetFrag. *Analytical and bioanalytical chemistry*, 1-18.

Ruttkies, C., Schymanski, E. L., Wolf, S., Hollender, J., & Neumann, S. (2016). MetFrag relaunched: incorporating strategies beyond in silico fragmentation. *Journal of cheminformatics*, *8*(1), 3.

Saad, O. M., & Leary, J. A. (2005). Heparin sequencing using enzymatic digestion and ESI-MS n with HOST: a heparin/HS oligosaccharide sequencing tool. *Analytical chemistry*, *77*(18), 5902-5911.

Sabater, C., Corzo, N., Olano, A., & Montilla, A. (2018a). Enzymatic extraction of pectin from artichoke (Cynara scolymus L.) by-products using Celluclast® 1.5 L. *Carbohydrate polymers*, *190*, 43-49.

Sabater, C., Montilla, A., Ovejero, A., Prodanov, M., Olano, A., & Corzo, N. (2018b). Furosine and HMF determination in prebiotic-supplemented infant formula from Spanish market. *Journal of Food Composition and Analysis*, *66*, 65–73.

Sabater, C., Olano, A., Corzo, N., & Montilla, A. (2019a). GC–MS characterization of novel artichoke (Cynara scolymus) pectic-oligosaccharides mixtures by the application of machine learning algorithms and competitive fragmentation modelling. *Carbohydrate polymers*, *205*, 513-523.

Sabater, C., Ferreira-Lazarte, A., Montilla, A., & Corzo, N. (2019b). Enzymatic Production and Characterization of Pectic Oligosaccharides Derived from Citrus and Apple Pectins: A GC-MS Study Using Random Forests and Association Rule Learning *Journal of Agricultural and Food Chemistry*, *67*, 7435-7447.

Sabater, C., Sabater, V., Olano, A., Montilla, A., & Corzo, N. (2020). Ultrasound-assisted extraction of pectin from artichoke by-products. An artificial neural network approach to pectin characterization. *Food Hydrocolloids*, *98*, 105238.

Sahasrabudhe, N. M., Beukema, M., Tian, L., Troost, B., Scholte, J., Bruininx, E., Bruggeman, G. van der Berg, M., Scheurink, A., Schols, H. A., Faas, M. M., & de Vos, P. (2018). Dietary fiber pectin directly blocks Toll-like receptor 2–1 and prevents doxorubicin-induced ileitis. *Frontiers in immunology*, *9*, 383.

Sałata, A., & Gruszecki, R. (2010). The quantitative analysis of poliphenolic compounds in different parts of the artichoke (Cynara scolymus L.) depending of growth stage of plants. *Acta Sci. Pol., Hortorum Cultus*, *9*(3), 175-181.

Sanders, M. E., Merenstein, D. J., Reid, G., Gibson, G. R., & Rastall, R. A. (2019). Probiotics and prebiotics in intestinal health and disease: from biology to the clinic. *Nature Reviews Gastroenterology & Hepatology*, 16, 605-616.

Santo Domingo, C., Rojas, A. M., Fissore, E. N., & Gerschenson, L. N. (2019). Rheological behavior of soluble dietary fiber fractions isolated from artichoke residues. *European Food Research and Technology*, *245*(6), 1239-1249.

Santos, J. D. G., Espeleta, A. F., Branco, A., & de Assis, S. A. (2013). Aqueous extraction of pectin from sisal waste. *Carbohydrate Polymers*, *92*(2), 1997-2001.

Satari, B., & Karimi, K. (2018). Citrus processing wastes: environmental impacts, recent advances, and future perspectives in total valorization. *Resources, Conservation and Recycling*, *129*, 153-167.

Schütz, K., Muks, E., Carle, R., & Schieber, A. (2006). Separation and quantification of inulin in selected artichoke (Cynara scolymus L.) cultivars and dandelion (Taraxacum officinale WEB. ex WIGG.) roots by high-performance anion exchange chromatography with pulsed amperometric detection. *Biomedical Chromatography*, *20*(12), 1295-1303.

Schymanski, E. L., Ruttkies, C., Krauss, M., Brouard, C., Kind, T., Dührkop, K., Allen, F., Vaniya, A., Verdegem, D., Böcker, S., Rousu, J., Shen, H., Tsugawa, H., Sajed, T., Fiehn, O., Ghesquière, B., & Neumann, S. (2017). Critical Assessment of Small Molecule Identification 2016: automated methods: *Journal of Cheminformatics*, 9, 1-21.

Silva, D., Martins, E. D. S., Silva, R. D., & Gomes, E. (2002). Pectinase production by Penicillium viridicatum RFC3 by solid state fermentation using agricultural wastes and agro-industrial by-products. *Brazilian Journal of Microbiology*, *33*(4), 318-324.

Silveira, A. L. M., Ferreira, A. V. M., de Oliveira, M. C., Rachid, M. A., da Cunha Sousa, L. F., dos Santos Martins, F., Gomes-Santos, A. C., Vieira, A. T., & Teixeira, M. M. (2017). Preventive rather than therapeutic treatment with high fiber diet attenuates clinical and inflammatory markers of acute and chronic DSS-induced colitis in mice. *European journal of nutrition*, *56*(1), 179-191.

Singthong, J., Cui, S. W., Ningsanond, S., & Goff, H. D. (2004). Structural characterization, degree of esterification and some gelling properties of Krueo Ma Noy (Cissampelos pareira) pectin. *Carbohydrate polymers*, *58*(4), 391-400.

Skarysz, A., Alkhalifah, Y., Darnley, K., Eddleston, M., Hu, Y., McLaren, D. B., Nailon, W.H., Salmanz, D., Sykora, M., Thomas, C.L.P., & Soltoggio, A. (2018, July). Convolutional neural networks for automated targeted analysis of raw gas

chromatography-mass spectrometry data. In *2018 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.

Song, Y. R., Sung, S. K., Jang, M., Lim, T. G., Cho, C. W., Han, C. J., & Hong, H. D. (2018). Enzyme-assisted extraction, chemical characteristics, and immunostimulatory activity of polysaccharides from Korean ginseng (Panax ginseng Meyer). *International journal of biological macromolecules*, *116*, 1089-1097.

Soria, A. C., Villamiel, M., & Montilla, A. (2017). Ultrasound Effects on Processes and Reactions Involving Carbohydrates. *Ultrasound in Food Processing: Recent Advances*, UK: John Wiley & Sons, Ltd, Chichester, 437–463.

Soria, A. C.; Rodriguez-Sanchez, S.; Sanz, J.; Martinez-Castro, I. Gas Chromatographic Analysis of Food Bioactive Oligosaccharides. *Food Oligosaccharides. Production, Analysis and Bioactivity*, F. J. Moreno, M. L. Sanz (Eds.), John Wiley & Sons, Ltd, Chichester, UK, 2014, pp. 370–398.

Stastny, K., Stepanova, H., Hlavova, K., & Faldyna, M. (2019). Identification and determination of deoxynivalenol (DON) and deepoxy-deoxynivalenol (DOM-1) in pig colostrum and serum using liquid chromatography in combination with high resolution mass spectrometry (LC-MS/MS (HR)). *Journal of Chromatography B*, *1126*, 121735.

Sun, L., Ropartz, D., Cui, L., Shi, H., Ralet, M. C., & Zhou, Y. (2019). Structural characterization of rhamnogalacturonan domains from Panax ginseng CA Meyer. *Carbohydrate polymers*, *203*, 119-127.

Sun, Y., Fan, L., Mian, W., Zhang, F., Liu, X., Tang, Y., Zeng, X., Mei, Q., & Li, Y. (2018). Modified apple polysaccharide influences MUC-1 expression to prevent

ICR mice from colitis-associated carcinogenesis. *International journal of biological macromolecules*, *120*, 1387-1395.

Swennen, K., Courtin, C. M., & Delcour, J. A. (2006). Non-digestible oligosaccharides with prebiotic properties. *Critical reviews in food science and nutrition*, *46*(6), 459-471.

Szabó, O. E., & Csiszár, E. (2013). The effect of low-frequency ultrasound on the activity and efficiency of a commercial cellulase enzyme. *Carbohydrate polymers*, *98*(2), 1483-1489.

Tamiello, C. S., do Nascimento, G. E., Iacomini, M., & Cordeiro, L. M. (2018). Arabinogalactan from edible jambo fruit induces different responses on cytokine secretion by THP-1 macrophages in the absence and presence of proinflammatory stimulus. *International journal of biological macromolecules*, *107*, 35-41.

Tan, H., Chen, W., Liu, Q., Yang, G., & Li, K. (2018). Pectin Oligosaccharides Ameliorate Colon Cancer by Regulating Oxidative Stress-and Inflammation-Activated Signaling Pathways. *Frontiers in immunology*, *9*, 1504.

Tao, J., Huang, J., Yu, L., Li, Z., Liu, H., Yuan, B., & Zeng, D. (2018). A new methodology combining microscopy observation with Artificial Neural Networks for the study of starch gelatinization. *Food hydrocolloids*, *74*, 151-158.

Tao, Y., Wu, D., Zhang, Q. A., & Sun, D. W. (2014). Ultrasound-assisted extraction of phenolics from wine lees: Modeling, optimization and stability of extracts during storage. *Ultrasonics sonochemistry*, *21*(2), 706-715.

Tapre, A. R., & Jain, R. K. (2014). Pectinases: Enzymes for fruit processing industry. *International Food Research Journal*, 21(2), 447-453.

Terkmane, N., Krea, M., & Moulai-Mostefa, N. (2016). Optimization of inulin extraction from globe artichoke (Cynara cardunculus L. subsp. scolymus (L.) Hegi.) by electromagnetic induction heating process. *International Journal of Food Science & Technology*, *51*(9), 1997-2008.

Tingirikari, J. M. R. (2018). Microbiota-accessible pectic poly-and oligosaccharides in gut health. *Food & function*, *9*(10), 5059-5073.

Toba, T., Watanabe, A., & Adachi, S. (1983). Quantitative changes in sugars, especially oligosaccharides, during fermentation and storage of yogurt. *Journal of Dairy Science*, *66*(1), 17-20.

Tuszynski, J. (2014). caTools: Tools: moving window statistics, GIF, Base64, ROC AUC, etc.. *R package version 1.17.1*. URL https://CRAN.R-project.org/package=caTools Last accessed: 30/07/18

Uarrota, V. G., Moresco, R., Coelho, B., da Costa Nunes, E., Peruch, L. A. M., de Oliveira Neubert, E., Rocha, M., & Maraschin, M. (2014). Metabolomics combined with chemometric tools (PCA, HCA, PLS-DA and SVM) for screening cassava (*Manihot esculenta* Crantz) roots during postharvest physiological deterioration. *Food Chemistry*, *161*, 67-78.

Ueno, K., Koga, T., Kato, K., Golenbock, D. T., Gendler, S. J., Kai, H., & Kim, K. C. (2008). MUC1 mucin is a negative regulator of toll-like receptor signaling. *American journal of respiratory cell and molecular biology*, *38*(3), 263-268.

USDA. (2017). United States Department of Agriculture. Agricultural Research Service USDA Food Composition Databases. URL https://ndb.nal.usda.gov/ndb/search/list Last accessed: 30/07/18

Uzuner, S., & Cekmecelioglu, D. (2015). Enhanced pectinase production by optimizing fermentation conditions of Bacillus subtilis growing on hazelnut shell hydrolyzate. *Journal of Molecular Catalysis B: Enzymatic*, *113*, 62-67.

Van Laere, K. M., Hartemink, R., Bosveld, M., Schols, H. A., & Voragen, A. G. (2000). Fermentation of plant cell wall derived polysaccharides and their corresponding oligosaccharides by intestinal bacteria. *Journal of Agricultural and Food Chemistry*, *48*(5), 1644-1652.

Varmuza, K., He, P., & Fang, K. T. (2003). Boosting applied to classification of mass spectral data. *Journal of Data Science*, *1*(4), 391-404.

Villamiel, M., Montilla, A., Olano, A., & Corzo, N. (2014). Production and bioactivity of oligosaccharides derived from lactose. *Food oligosaccharides: production, analysis and bioactivity*, 137-167.

Villanueva-Suárez, M. J., Mateos-Aparicio, I., Pérez-Cózar, M. L., Yokoyama, W., & Redondo-Cuenca, A. (2019). Hypolipidemic effects of dietary fibre from an artichoke by-product in Syrian hamsters. *Journal of Functional Foods*, *56*, 156-162.

Vogt, J. A., Wachter, U., & Georgieff, M. (2003). Non-linearity in the quadrupole detector system: implications for the determination of the 13C mass distribution of an ion fragment. *Journal of mass spectrometry*, *38*(2), 222-230.

Walters, W. A., Xu, Z., & Knight, R. (2014). Meta-analyses of human gut microbes associated with obesity and IBD. *FEBS letters*, *588*(22), 4223-4233.

Walton, G. E., van den Heuvel, E. G., Kosters, M. H., Rastall, R. A., Tuohy, K. M., & Gibson, G. R. (2012). A randomized crossover study investigating the effects of galacto-oligosaccharides on the faecal microbiota in men and women over 50 years of age. *British Journal of Nutrition*, *107*(10), 1466-1475.

Wang, W., Wu, X., Chantapakul, T., Wang, D., Zhang, S., Ma, X., Ding, T., Ye, X. Q., & Liu, D. (2017). Acoustic cavitation assisted extraction of pectin from waste grapefruit peels: A green two-stage approach and its general mechanism. *Food Research International*, *102*, 101-110.

Wang, W., Ma, X., Jiang, P., Hu, L., Zhi, Z., Chen, J. Ding, T., Ye, X., & Liu, D. (2016). Characterization of pectin from grapefruit peel: A comparison of ultrasound-assisted and conventional heating extractions. *Food Hydrocolloids*, *61*, 730-739.

Wang, W., Ma, X., Xu, Y., Cao, Y., Jiang, Z., Ding, T., Ye, X., & Liu, D. (2015). Ultrasound-assisted heating extraction of pectin from grapefruit peel: Optimization and comparison with the conventional method. *Food Chemistry*, *178*, 106-114.

Wang, X., Chen, Q., & Lü, X. (2014a). Pectin extracted from apple pomace and citrus peel by subcritical water. *Food Hydrocolloids*, *38*, 129-137.

Wang, J., Sun, B., Liu, Y., & Zhang, H. (2014b). Optimization of ultrasound-assisted enzymatic extraction of arabinoxylan from wheat bran. *Food chemistry*, *150*, 482-488.

Wefers, D., & Bunzel, M. (2016). Arabinan and galactan oligosaccharide profiling by high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD). *Journal of agricultural and food chemistry*, *64*(22), 4656-4664.

Westerbeek, E. A., van den Berg, J. P., Lafeber, H. N., Fetter, W. P., Boehm, G., Twisk, J. W., & van Elburg, R. M. (2009). Neutral and acidic oligosaccharides in preterm infants: a randomized, double-blind, placebo-controlled trial. *The American journal of clinical nutrition*, *91*(3), 679-686.

Westereng, B., Coenen, G. J., Michaelsen, T. E., Voragen, A. G., Samuelsen, A. B., Schols, H. A., & Knutsen, S. H. (2009). Release and characterization of single side chains of white cabbage pectin and their complement-fixing activity. *Molecular nutrition & food research*, *53*(6), 780-789.

Wikiera, A., Mika, M., Starzyńska-Janiszewska, A., & Stodolak, B. (2015a). Development of complete hydrolysis of pectins from apple pomace. *Food Chemistry*, *172*, 675-680.

Wikiera, A., Mika, M., Starzyńska-Janiszewska, A., & Stodolak, B. (2015b). Application of Celluclast 1.5 L in apple pectin extraction. *Carbohydrate Polymers*, *134*, 251-257.

Xia, J. M., Wu, X. J., & Yuan, Y. J. (2007). Integration of wavelet transform with PCA and ANN for metabolomics data-mining. *Metabolomics*, *3*(4), 531-537.

Xu, J. K., Li, M. F., & Sun, R. C. (2015). Identifying the impact of ultrasound-assisted extraction on polysaccharides and natural antioxidants from Eucommia ulmoides Oliver. *Process Biochemistry*, *50*(3), 473-481.

Xu, M., Wang, J., & Zhu, L. (2019). The qualitative and quantitative assessment of tea quality based on E-nose, E-tongue and E-eye combined with chemometrics. *Food chemistry*, *289*, 482-489.

Yang, Q., Shi, L., Han, J., Zha, Y., & Zhu, P. (2019). Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. *Field Crops Research*, *235*, 142-153.

Yang, J. S., Mu, T. H., & Ma, M. M. (2018a). Extraction, structure, and emulsifying properties of pectin from potato pulp. *Food chemistry*, *244*, 197-205.

Yang, Y., Wang, Z., Hu, D., Xiao, K., & Wu, J. Y. (2018b). Efficient extraction of pectin from sisal waste by combined enzymatic and ultrasonic process. *Food hydrocolloids*, *79*, 189-196.

Yapo, B. M. (2009). Biochemical characteristics and gelling capacity of pectin from yellow passion fruit rind as affected by acid extractant nature. *Journal of agricultural and food chemistry*, *57*(4), 1572-1578.

Yapo, B. M., & Koffi, K. L. (2008). Dietary Fiber Components in Yellow Passion Fruit Rind: A Potential Fiber Source. *Journal of Agricultural and Food Chemistry*, *56*(14), 5880-5883.

Yasui, Y., Pepe, M., Thompson, M. L., Adam, B. L., Wright Jr, G. L., Qu, Y., Potter, J. D., Winget, M., Thornquist, M., & Feng, Z. (2003). A data-analytic strategy for protein biomarker discovery: profiling of high-dimensional proteomic data for cancer detection. *Biostatistics*, *4*, 449-463.

Yi, L., Dong, N., Yun, Y., Deng, B., Ren, D., Liu, S., & Liang, Y. (2016). Chemometric methods in data processing of mass spectrometry-based metabolomics: A review. *Analytica chimica acta*, *914*, 17-34.

You, H., Kim, H., Joo, D. K., Lee, S. M., Kim, J., & Choi, S. (2019, February). Classification of Food Powders with Open Set using Portable VIS-NIR Spectrometer. In *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)* (pp. 423-426). IEEE.

Yu, M., Olkowicz, M., & Pawliszyn, J. (2019). Structure/reaction directed analysis for LC-MS based untargeted analysis. *Analytica chimica acta*, *1050*, 16-24.

Yu, P., & Xu, C. (2018). Production optimization, purification and characterization of a heat-tolerant acidic pectinase from Bacillus sp. ZJ1407. *International journal of biological macromolecules*, *108*, 972-980.

Yuliarti, O., Goh, K. K., Matia-Merino, L., Mawson, J., & Brennan, C. (2015). Extraction and characterization of pomace pectin from gold kiwifruit (Actinidia chinensis). *Food chemistry*, *187*, 290-296.

Zhang, Y., Pan, X., Ran, S., & Wang, K. (2019). Purification, structural elucidation and anti-inflammatory activity in vitro of polysaccharides from Smilax china L. *International journal of biological macromolecules*, *139*, 233-243.

Zheng, Y., Li, L., Feng, Z., Wang, H., Mayo, K. H., Zhou, Y., & Tai, G. (2018). Preparation of individual galactan oligomers, their prebiotic effects, and use in estimating galactan chain length in pectin-derived polysaccharides. *Carbohydrate polymers*, *199*, 526-533.

Zhu, R., Wang, C., Zhang, L., Wang, Y., Chen, G., Fan, J., Jia, Y., Yan, F., & Ning, C. (2019). Pectin oligosaccharides from fruit of Actinidia arguta: Structure-activity relationship of prebiotic and antiglycation potentials. *Carbohydrate polymers*, *217*, 90-97.

Zhu, G., Wang, H., Wang, T., & Shi, F. (2017). Ginsenoside Rg1 attenuates the inflammatory response in DSS-induced mice colitis. *International immunopharmacology*, *50*, 1-5.

Zitnik, M., Nguyen, F., Wang, B., Leskovec, J., Goldenberg, A., & Hoffman, M. M. (2019). Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. *Information Fusion*, *50*, 71-91.

Annex I:

Supplementary material

Annex I.I:

Supplementary material of Article I

11. ANNEX I

11.1. Annex I.I: Supplementary material of Article I

Table S1.1. Determination of the pectinase and cellulase activities of different commercial enzyme preparations.

	Pectinase activity (U mL^{-1})	Cellulase activity (U mL^{-1})
Celluclast®1.5	-	111.0 ± 6.4
Pectinex Ultra SP-L®	543 ± 20	1.4 ± 0.1
Pectinase (*A. niger*)	532 ± 7	0.7 ± 0.0
Viscozyme® L	3778 ± 67	0.7 ± 0.1
Viscozyme® Barley HT	-	63.9 ± 4.2
Carezyme® 4500 L	-	5.0 ± 0.1
Celullase (*Trichoderma viridae)***	-	70.1 ± 3.7**
Cellulase (*Aspergillus niger)***	1712 ± 10	242.4 ± 13.2

**U g^{-1}

Table S1.2. Monomeric composition of extracted pectin from artichoke by-product powder (mg g^{-1}).Data obtained by acid hydrolysis with TFA 2M at 110 °C under inert conditions for 4 h.

	*Monomeric composition (mg g^{-1})**						
Run	**Arabinose**	**Rhamnose**	**Fructose**	**Galactose**	**Mannose**	**Glucose**	**Galacturonic acid**
1	21.6 ± 1.6	5.1 ± 0.3	0.5 ± 0.0	5.9 ± 0.8	5.5 ± 0.2	9.9 ± 0.7	49.8 ± 2.4
2	14.3 ± 0.1	3.3 ± 0.1	0.3 ± 0.0	3.6 ± 0.1	4.4 ± 0.3	6.8 ± 0.3	47.8 ± 0.1
3	19.5 ± 0.0	4.6 ± 0.1	0.4 ± 0.0	4.8 ± 0.3	4.5 ± 0.0	10.1 ± 0.3	52.2 ± 0.9
4	8.7 ± 0.1	2.0 ± 0.0	0.2 ± 0.0	2.4 ± 0.0	2.9 ± 0.0	4.0 ± 0.2	30.5 ± 0.5
5	21.4 ± 1.4	5.0 ± 0.2	0.4 ± 0.0	4.8 ± 0.1	4.7 ± 0.4	10.4 ± 1.3	58.1 ± 5.1
6	20.5 ± 0.1	4.6 ± 0.0	0.8 ± 0.3	4.7 ± 0.5	5.8 ± 0.4	9.6 ± 0.5	50.6 ± 1.2
7	15.2 ± 0.2	3.6 ± 0.1	0.2 ± 0.0	3.7 ± 0.2	3.5 ± 0.2	6.2 ± 0.3	45.3 ± 0.1
8	20.9 ± 1.4	4.8 ± 0.3	0.4 ± 0.0	4.8 ± 0.3	5.7 ± 1.0	10.3 ± 2.3	51.6 ± 2.1
9	25.3 ± 0.4	5.7 ± 0.0	0.7 ± 0.0	6.3 ± 0.3	7.1 ± 0.5	13.8 ± 0.8	57.6 ± 1.9
10	8.4 ± 0.7	1.9 ± 0.0	0.2 ± 0.0	2.0 ± 0.0	3.4 ± 0.3	4.7 ± 0.2	32.2 ± 3.1
11	24.3 ± 1.6	5.8 ± 0.7	0.4 ± 0.1	6.1 ± 0.5	5.1 ± 0.5	7.1 ± 1.1	52.1 ± 3.3
12	21.9 ± 0.1	4.9 ± 0.0	1.0 ± 0.0	6.0 ± 0.0	6.5 ± 0.3	14.0 ± 0.8	48.7 ± 1.5
13	13.8 ± 0.0	3.3 ± 0.1	0.2 ± 0.0	3.3 ± 0.3	3.6 ± 1.0	5.9 ± 0.1	37.4 ± 0.4
14	16.0 ± 0.9	3.6 ± 0.2	0.2 ± 0.0	3.8 ± 0.0	5.5 ± 0.1	7.7 ± 0.3	44.6 ± 0.7
15	16.8 ± 0.7	3.7 ± 0.2	0.8 ± 0.1	4.1 ± 0.1	5.9 ± 0.2	7.2 ± 0.5	38.6 ± 1.3
16	22.8 ± 0.5	5.2 ± 0.1	0.5 ± 0.0	5.5 ± 0.2	5.5 ± 0.1	5.2 ± 0.8	51.6 ± 3.4
17	15.5 ± 1.0	3.6 ± 0.2	0.5 ± 0.0	4.3 ± 0.1	4.0 ± 0.2	8.1 ± 0.7	41.9 ± 3.4

*Values corrected according to the degradation coefficients of the monosaccharides standards hydrolysed under the same conditions.

Annex I.II:

Supplementary material of Article II

11.2. Annex I.II: Supplementary material of Article II

Table S2.1. Importance coefficients obtained in the multilayer perceptron (MLP) classification of artichoke pectin extracted with different methods according to their process yield and composition parameters (Model I). Importance coefficients were determined by the sum of the product of raw input-hidden and hidden-output connection weights. GalA: galacturonic acid, Xyl: xylose; Ara: arabinose, Rha: rhamnose, Fru: fructose, Gal: galactose, Man: mannose, Glc: glucose.

	Importance coefficients				
Parameters	**Sodium citrate**	**Nitric acid**	**Ultrasounds**	**Ultrasounds + Enzyme**	**Enzyme**
Yield (%)	-3.26	-1.94	-6.87	0.57	6.53
Average Mw (kDa)	2.47	1.03	-2.81	-1.48	-0.83
GalA (%)	0.82	3.25	-2.35	-0.80	0.16
Xyl (%)	-2.59	-1.24	-5.74	7.34	1.46
Ara (%)	-0.79	-3.93	1.32	2.27	-0.54
Rha (%)	-2.11	2.41	-0.20	-0.08	2.40
Fru (%)	-5.72	2.00	1.93	4.04	1.85
Gal (%)	-2.17	-0.29	-0.68	3.12	0.72
Man (%)	-2.61	0.12	-0.51	0.60	2.36
Glc (%)	1.00	-5.21	7.90	0.60	-5.03
Degree of branching (Rha/GalA)	-0.77	-0.31	1.43	1.58	-0.88
Linearity pectin backbone [GalA/(Rha+Ara+Gal)]	-1.19	3.81	-1.60	-1.08	2.00
Extent of branching [(Ara+Gal)/Rha]	-0.84	-5.13	2.37	2.16	-0.59
Unidentified Mn (mg g^{-1})	-2.16	1.47	-5.80	-0.82	5.53
Degree of methyl-esterification (%)	2.96	-0.10	-0.07	-0.41	-3.22

Table S2.2. Most influential wavenumbers (cm^{-1}) in the multilayer perceptron (MLP) classification of artichoke pectin extracted with different methods according to their FT-IR spectra (Model II). Importance coefficients were determined by the sum of the product of raw input-hidden and hidden-output connection weights.

Sodium citrate		**Nitric acid**		**Ultrasounds**		**Ultrasounds + Enzyme**		**Enzyme**	
***cm*$^{-1}$**	**Importance**	***cm*$^{-1}$**	**Importance**	***cm*$^{-1}$**	**Importance**	***cm*$^{-1}$**	**Importance**	***cm*$^{-1}$**	**Importance**
1042	100.0	**1771**	100.0	**1776**	100.0	**2786**	100.0	**1831**	100.0
1055	84.0	**1063**	95.5	**1777**	95.8	**1308**	82.4	**1081**	93.8
1817	81.8	**1718**	93.1	**1778**	92.4	**1145**	80.1	**2772**	91.8
1684	78.8	**1762**	92.9	**1825**	89.9	**1745**	74.8	**1830**	89.9
1044	75.4	**1000**	84.7	**1104**	88.1	**1142**	74.3	**1776**	89.1
1258	73.1	**1154**	84.0	**1773**	87.2	**1230**	71.9	**1131**	77.4
1271	73.0	**2822**	77.3	**2703**	69.5	**1723**	69.7	**1019**	74.4
1024	73.0	**2805**	75.0	**1774**	68.2	**2759**	68.6	**2748**	74.2
1083	72.7	**1769**	73.8	**2723**	66.4	**2828**	67.0	**1838**	73.0
1700	72.4	**1103**	72.7	**1772**	62.6	**1680**	63.3	**2800**	71.9
1764	68.4	**1015**	72.4	**1193**	62.6	**2892**	62.0	**1128**	68.8
1052	68.0	**1730**	70.6	**1203**	61.1	**1187**	61.1	**2810**	68.7
1063	67.8	**1135**	69.3	**2827**	61.1	**1721**	59.5	**1040**	67.5
1151	65.0	**2877**	68.8	**1106**	59.4	**1713**	59.0	**2736**	64.8
1134	63.9	**1152**	65.9	**1082**	58.3	**1125**	58.8	**1762**	64.6
1780	63.7	**1765**	65.4	**2898**	58.1	**2733**	58.4	**1036**	64.2
1129	63.7	**1011**	64.2	**1204**	57.3	**1132**	58.3	**1004**	63.8
1154	63.4	**1080**	63.1	**1239**	54.8	**1160**	58.2	**2816**	61.2
1020	59.8	**1840**	57.6	**2749**	52.3	**1300**	57.9	**2803**	60.8
1050	57.7	**1724**	56.8	**1271**	52.3	**1221**	57.7	**2745**	60.6
1135	57.0	**1255**	56.4	**2855**	50.8	**2795**	57.7	**1800**	60.0
1002	56.8	**1808**	56.4	**1771**	50.3	**2884**	57.3	**1715**	59.7
1812	54.8	**1760**	54.5	**2899**	47.5	**1727**	55.2	**1002**	59.5
1275	54.4	**1709**	53.3	**1197**	47.5	**2706**	54.7	**1035**	59.4
1681	53.4	**2803**	52.5	**1264**	47.4	**1220**	54.2	**1010**	59.2
2733	53.0	**1013**	52.4	**2780**	46.5	**1275**	52.7	**2757**	57.6
1113	52.9	**1684**	50.3	**1171**	46.5	**1292**	52.2	**2885**	57.5
2866	52.4	**1817**	50.1	**1190**	46.5	**1156**	52.1	**2853**	56.3
2868	52.3	**1703**	49.8	**2701**	46.3	**2753**	51.5	**1769**	54.8
1302	52.2	**2817**	49.2	**2760**	46.1	**1262**	50.7	**1104**	53.3

Table S2.3. Most influential *m/z* ions in the multilayer perceptron (MLP) classification of GC-MS spectra (n = 555) of identified (n = 8) and unidentified (n = 12) monosaccharides found in hydrolysates of artichoke pectins extracted with different methods (Model III). Importance coefficients were determined by the sum of the product of raw input-hidden and hidden-output connection weights.

Xylose		Arabinose		Rhamnose		Fructose		Galactose		Mannose	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
427	100.0	**177**	100.0	**337**	100.0	**150**	100.0	**337**	100.0	**195**	100.0
133	38.7	**222**	55.1	**279**	69.9	**132**	98.5	**338**	92.4	**265**	75.6
324	37.4	**382**	48.6	**119**	68.4	**165**	68.2	**134**	83.2	**221**	52.3
309	28.8	**426**	45.8	**397**	47.9	**297**	57.9	**193**	68.6	**134**	44.4
426	22.4	**134**	40.2	**338**	33.0	**133**	55.9	**177**	65.0	**324**	42.5
178	18.4	**324**	38.8	**323**	27.6	**220**	52.0	**207**	62.5	**367**	40.4
280	12.0	**235**	37.7	**369**	25.3	**438**	45.7	**222**	49.0	**133**	38.9
339	7.5	**150**	31.5	**280**	25.2	**468**	45.5	**397**	37.8	**309**	38.6
149	7.4	**280**	29.4	**382**	23.2	**235**	44.4	**223**	37.1	**339**	37.0
325	4.1	**251**	27.9	**221**	15.3	**238**	41.5	**208**	33.2	**208**	32.9
207	3.0	**207**	25.8	**237**	11.6	**166**	35.7	**163**	32.4	**177**	28.2
411	1.0	**279**	25.3	**355**	11.1	**351**	34.5	**150**	31.3	**149**	26.5
237	-2.0	**178**	21.5	**195**	10.7	**237**	32.6	**165**	29.7	**163**	25.4
468	-2.2	**208**	16.4	**150**	10.0	**540**	32.4	**382**	28.3	**311**	24.2
107	-4.7	**197**	16.2	**469**	9.0	**180**	20.2	**149**	28.0	**207**	22.5
120	-5.3	**325**	11.3	**427**	9.0	**397**	17.3	**295**	27.4	**295**	14.2
323	-6.7	**427**	10.6	**151**	7.0	**151**	17.3	**121**	27.2	**279**	7.5
197	-6.7	**163**	10.5	**120**	6.7	**177**	16.3	**323**	24.8	**369**	-3.7
221	-7.2	**425**	9.6	**207**	6.1	**193**	10.6	**135**	22.5	**280**	-5.6
355	-7.3	**149**	8.2	**295**	5.6	**134**	7.1	**133**	22.2	**355**	-7.2
251	-9.0	**121**	7.6	**468**	3.8	**338**	6.5	**251**	22.0	**426**	-8.1
222	-10.6	**311**	7.6	**165**	3.5	**369**	4.4	**369**	19.5	**251**	-8.1
235	-11.3	**223**	0.5	**325**	1.6	**223**	1.9	**221**	15.8	**209**	-19.3
425	-11.4	**135**	-1.3	**209**	0.7	**339**	1.4	**339**	14.0	**325**	-34.3
162	-28.0	**195**	-5.3	**121**	0.6	**427**	-0.1	**209**	13.0	**352**	-42.8
424	-39.0	**438**	-13.1	**135**	0.5	**295**	-1.4	**151**	13.0	**222**	-47.1
195	-52.8	**411**	-13.2	**339**	-0.2	**411**	-1.7	**178**	10.8	**148**	-62.1
223	-60.8	**166**	-24.1	**222**	-17.3	**118**	-4.0	**250**	9.5	**162**	-63.1
208	-79.3	**352**	-26.4	**309**	-30.7	**234**	-4.4	**120**	1.1.	**223**	-64.2
438	-99.0	**148**	-27.6	**148**	-41.9	**178**	-5.4	**235**	-1.2	**438**	-113.4

Table S2.3. Cont.

Glucose		Galacturonic acid		Unidentified 1		Unidentified 2		Unidentified 3	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
324	100.0	**265**	100.0	**279**	100.0	**151**	100.0	**177**	100.0
106	76.5	**426**	83.1	**309**	78.4	**469**	68.4	**178**	53.0
163	50.8	**425**	82.7	**133**	65.1	**193**	53.8	**469**	37.5
195	47.7	**295**	71.4	**411**	55.9	**149**	43.3	**251**	29.2
178	44.3	**223**	68.3	**353**	53.0	**195**	41.8	**197**	22.6
323	42.3	**427**	61.4	**469**	43.3	**355**	40.4	**163**	22.1
207	39.2	**221**	61.4	**351**	43.3	**107**	32.4	**150**	20.1
237	35.0	**351**	60.1	**265**	43.3	**223**	32.0	**468**	19.9
165	34.8	**337**	51.5	**339**	42.7	**411**	28.5	**355**	16.5
133	31.4	**280**	39.1	**223**	36.8	**353**	27.4	**222**	12.0
120	29.8	**338**	37.3	**369**	36.6	**339**	26.3	**353**	9.0
367	28.1	**367**	26.3	**135**	34.3	**251**	25.9	**209**	8.8
177	26.1	**382**	23.6	**355**	33.6	**133**	25.4	**382**	7.8
295	23.1	**222**	19.8	**468**	32.9	**468**	23.5	**325**	7.6
265	22.4	**397**	18.5	**222**	23.6	**165**	17.6	**351**	6.3
149	21.8	**411**	18.2	**209**	23.2	**351**	16.2	**339**	5.2
221	20.6	**209**	13.3	**338**	19.5	**209**	15.0	**324**	4.6
151	17.5	**353**	11.4	**107**	16.0	**325**	15.0	**297**	3.7
135	16.3	**207**	9.9	**337**	12.9	**369**	14.5	**120**	3.6
427	16.2	**369**	8.4	**251**	8.6	**197**	10.5	**235**	3.4
134	16.1	**324**	7.9	**221**	6.6	**280**	6.4	**369**	2.7
309	14.7	**193**	7.3	**197**	4.8	**150**	2.8	**134**	0.9
209	9.0	**135**	7.3	**280**	4.8	**279**	2.8	**223**	-16.9
208	7.9	**150**	5.6	**350**	-44.1	**207**	1.4	**337**	-25.5
280	2.2	**469**	3.5	**262**	-50.7	**323**	-1.3	**322**	-26.5
355	2.1	**468**	1.3	**224**	-52.9	**338**	-1.4	**160**	-43.5
411	-1.1	**339**	0.2	**397**	-59.5	**235**	-2.6	**411**	-82.9
224	-8.7	**250**	-2.4	**120**	-62.8	**166**	-3.3	**425**	-84.0
235	-13.1	**251**	-7.3	**180**	-81.6	**135**	-4.5	**426**	-88.2
382	-15.3	**160**	-8.7	**427**	-93.7	**220**	-10.1	**250**	-89.2

Table S2.3. Cont.

Unidentified 4		Unidentified 5		Unidentified 6		Unidentified 7		Unidentified 8	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
295	100.0	**367**	100.0	**297**	100.0	**424**	100.0	**337**	100.0
469	98.9	**265**	98.6	**338**	70.8	**119**	99.1	**352**	72.0
197	77.8	**193**	65.2	**367**	35.8	**118**	78.7	**336**	71.4
121	73.6	**297**	36.2	**222**	35.2	**297**	78.5	**311**	68.4
297	61.8	**338**	33.7	**149**	30.7	**325**	77.4	**133**	63.7
325	56.8	**369**	31.0	**235**	29.7	**355**	63.7	**121**	61.9
120	56.2	**221**	25.1	**311**	26.9	**150**	49.6	**350**	47.5
468	50.3	**209**	22.9	**208**	24.9	**224**	49.1	**208**	47.0
207	48.2	**469**	22.6	**382**	18.5	**353**	49.0	**248**	45.7
311	40.9	**251**	19.0	**309**	17.3	**121**	43.8	**160**	43.9
337	40.6	**133**	16.9	**237**	15.4	**132**	37.5	**282**	40.6
222	39.8	**397**	15.9	**163**	14.9	**262**	36.3	**353**	40.6
107	31.3	**149**	11.2	**425**	14.0	**338**	35.2	**180**	40.4
209	31.1	**208**	10.5	**411**	13.9	**222**	33.0	**355**	39.8
237	30.9	**237**	10.4	**107**	12.4	**468**	33.0	**151**	38.6
280	24.3	**411**	7.0	**193**	11.6	**351**	28.7	**250**	37.8
411	22.4	**280**	6.9	**207**	3.1	**135**	27.6	**207**	34.5
323	17.5	**324**	5.3	**178**	2.7	**162**	23.4	**197**	34.3
369	14.6	**337**	4.5	**197**	2.0	**280**	20.4	**209**	31.9
119	11.9	**120**	2.9	**339**	1.5	**250**	19.2	**262**	30.5
324	10.2	**425**	2.9	**121**	1.4	**134**	19.2	**309**	28.7
150	9.9	**107**	2.4	**251**	-0.6	**221**	17.9	**295**	27.2
251	8.7	**222**	2.3	**324**	-3.6	**120**	17.7	**193**	24.5
355	3.8	**355**	2.3	**151**	-5.1	**223**	17.3	**224**	23.7
351	3.7	**382**	-2.3	**468**	-44.4	**248**	15.3	**251**	23.5
322	-8.5	**235**	-17.6	**265**	-56.8	**469**	14.0	**222**	23.0
438	-14.6	**207**	-26.9	**162**	-80.6	**411**	13.5	**107**	22.8
382	-38.2	**468**	-35.3	**337**	-94.0	**151**	12.6	**351**	22.0
224	-49.5	**223**	-40.0	**223**	-101.2	**309**	12.4	**234**	20.3
426	-50.6	**197**	-45.8	**118**	-132.2	**180**	9.3	**134**	20.2

Table S2.3. Cont.

Unidentified 9		Unidentified 10		Unidentified 11		Unidentified 12	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
339	100.0	**311**	100.0	**133**	100.0	**193**	100.0
222	84.4	**325**	56.6	**297**	48.0	**208**	90.8
468	83.5	**120**	55.6	**311**	41.1	**135**	72.9
469	56.7	**411**	54.0	**107**	40.7	**209**	59.8
411	53.1	**197**	53.5	**121**	40.6	**295**	48.2
338	52.4	**195**	51.5	**134**	35.6	**411**	47.3
280	52.0	**324**	49.2	**119**	35.1	**223**	39.5
207	49.2	**237**	40.3	**251**	27.4	**237**	33.4
197	43.6	**280**	25.3	**105**	20.0	**311**	32.5
382	42.5	**165**	20.1	**135**	16.3	**251**	29.0
337	33.9	**221**	10.5	**207**	14.9	**221**	26.2
235	26.9	**323**	3.6	**324**	13.9	**120**	22.5
221	25.2	**251**	-0.4	**235**	9.1	**325**	22.4
135	23.2	**208**	-6.9	**411**	8.4	**355**	22.1
120	21.2	**121**	-10.2	**197**	6.9	**165**	19.1
150	18.2	**235**	-10.4	**209**	6.5	**207**	17.8
223	18.0	**297**	-11.9	**208**	6.5	**134**	17.2
369	16.1	**107**	-14.3	**149**	5.7	**197**	13.9
209	15.0	**209**	-16.1	**339**	4.6	**297**	11.6
425	14.6	**149**	-23.7	**165**	0.5	**107**	11.1
265	8.8	**355**	-23.8	**355**	0.4	**382**	9.1
107	8.2	**133**	-30.3	**237**	-0.1	**369**	6.8
178	7.6	**322**	-42.7	**223**	-2.2	**121**	5.9
351	7.5	**540**	-43.8	**427**	-6.7	**426**	2.2
193	7.3	**438**	-98.3	**160**	-22.4	**279**	5.5
163	-8.4	**282**	-108.0	**262**	-32.5	**163**	-28.6
282	-13.0	**337**	-128.3	**426**	-51.5	**336**	-56.1
248	-50.8	**250**	-173.1	**178**	-81.8	**119**	-78.1
397	-55.1	**162**	-183.8	**382**	-87.5	**425**	-108.9
540	-111.9	**350**	-194.5	**350**	-105.4	**132**	-113.3

Table S2.4. Sensitivity, specificity, true positives and negatives and balanced accuracy rates (%) for the artificial neural network (MLP) used to classify monomeric compounds present in pectin hydrolysates (Model III). Ara: arabinose, Gal: galactose, GalA: galacturonic acid, Glc: glucose, Man: mannose, Rha: rhamnose, Xyl: xylose, U: unidentified monomeric compound.

Parameter	Ara	Gal	GalA	Fru	Glc	Man	Rha	Xyl	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12
Sensitivity	100	81.8	100	100	100	100	94.1	100	100	100	100	100	100	100	80.0	100	100	100	100	100
Specificity	100	100	100	100	100	100	99.3	100	100	100	100	100	100	98.1	100	100	100	100	99.4	100
True positives	100	100	100	100	100	100	94.1	100	100	100	100	100	100	62.5	100	100	100	100	66.7	100
True negatives	100	98.7	100	100	100	100	99.3	100	100	100	100	100	100	100	98.7	100	100	100	100	100
Balanced Accuracy	100	90.9	100	100	100	100	96.7	100	100	100	100	100	100	99.1	90.0	100	100	100	99.7	100

Table S2.5. Sensitivity, specificity, true positives and negatives and balanced accuracy rates (%) for the naïve bayes (NB) model used to classify monomeric compounds present in pectin hydrolysates (Model III). Ara: arabinose, Gal: galactose, GalA: galacturonic acid, Glc: glucose, Man: mannose, Rha: rhamnose, Xyl: xylose, U: unidentified monomeric compound.

Parameter	Ara	Gal	GalA	Fru	Glc	Man	Rha	Xyl	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12
Sensitivity	81.8	100	100	100	100	90.0	100	100	100	100	100	100	100	100	80.0	88.9	-	100	75.0	100
Specificity	100	100	99.3	100	100	100	100	98.8	100	100	100	100	100	98.7	100	100	99.4	100	100	99.4
True positives	100	100	94.1	100	100	100	100	50.0	100	100	100	100	100	75.0	100	100	-	100	100	66.7
True negatives	98.7	100	100	100	100	99.4	100	100	100	100	100	100	100	100	98.7	99.4	-	100	99.4	100
Balanced Accuracy	90.9	100	99.7	100	100	95.0	100	99.4	100	100	100	100	100	99.4	90.0	94.4	-	100	87.5	99.7

Table S2.6. Sensitivity, specificity, true positives and negatives and balanced accuracy rates (%) for the generalized linear model elastic-net (GLMNET) model used to classify monomeric compounds present in pectin hydrolysates (Model III). Ara: arabinose, Gal: galactose, GalA: galacturonic acid, Glc: glucose, Man: mannose, Rha: rhamnose, Xyl: xylose, U: unidentified monomeric compound.

Parameter	**Ara**	**Gal**	**GalA**	**Fru**	**Glc**	**Man**	**Rha**	**Xyl**	**U1**	**U2**	**U3**	**U4**	**U5**	**U6**	**U7**	**U8**	**U9**	**U10**	**U11**	**U12**
Sensitivity	90.0	90.0	93.8	100	100	81.8	100	100	88.9	100	100	100	100	100	88.9	72,7	-	100	100	-
Specificity	100	100	98.7	100	100	100	99.3	99.4	100	99.4	100	100	100	99.4	100	100	99.4	100	99.4	98.2
True positives	100	100	88.2	100	100	100	94.1	75.0	100	87.5	100	100	100	87.5	100	100	-	100	66.7	-
True negatives	99.4	99.4	99.3	100	100	98.7	100	100	99.4	100	100	100	100	100	99.4	98.1	99.4	100	100	-
Balanced Accuracy	95.0	95.0	96.2	100	100	90.9	99.7	99.7	94.4	99.7	100	100	100	99.7	94.4	86.4	49.7	100	99.7	-

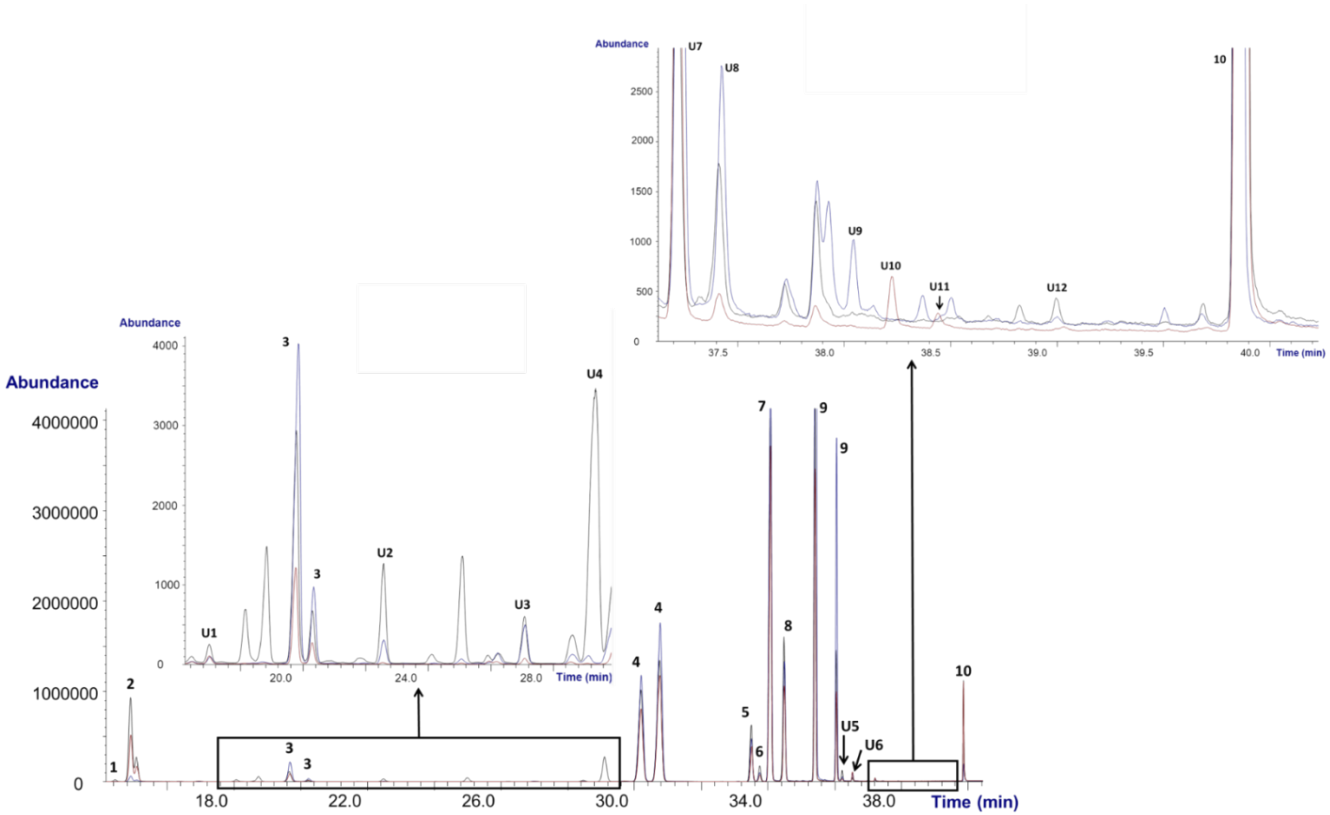


Figure S2.1. GC-MS profiles of identified (n= 8) and unidentified (U, n = 12) carbohydrates (released during Viscozyme hydrolysis) detected in hydrolysates of artichoke pectin extracted using different methods: US for 6 h (red), combination of US and Celluclast for 6 h (black) and nitric acid (blue). Peaks: 1) Xylose, 2) Arabinose, 3) Rhamnose, 4) Fructose (present in Viscozyme, only fructose released was quantified), 5) Galactose, 6) Mannose, 7) Glucose, 8) Galactose+Mannose+Glucose, 9) Galacturonic acid, 10: Internal standard, U: unknown monosaccharide.

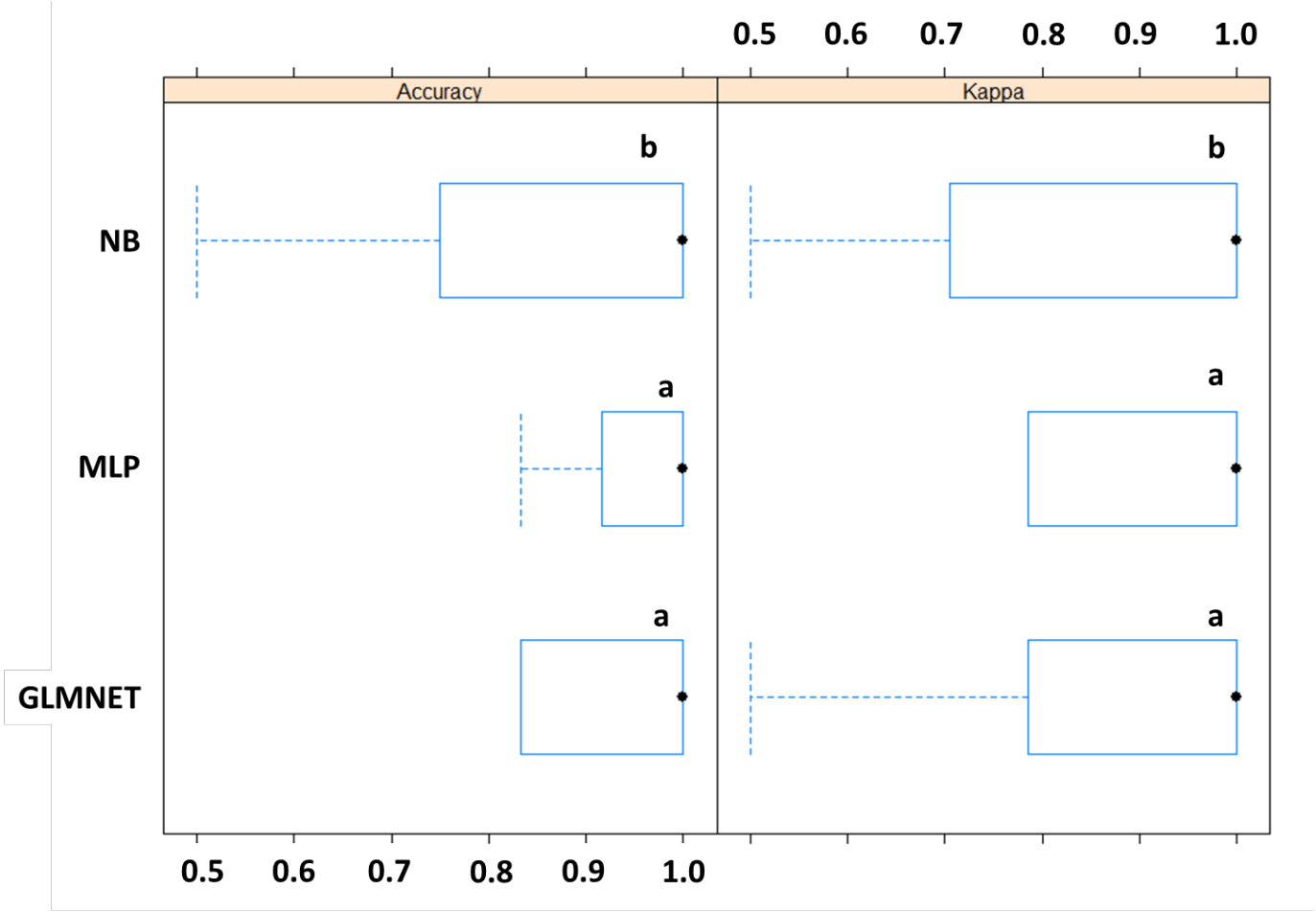


Figure S2.2. Comparative account of different three algorithms used for pectin classification using composition parameters (Model I): naïve bayes classifier (NB), multilayer perceptron (MLP, artificial neural network) and generalized linear model elastic-net (GLMNET). Differences between models were calculated via their resampling distributions (number of resamples = 10). [a,b] Statistically significant differences between algorithms.

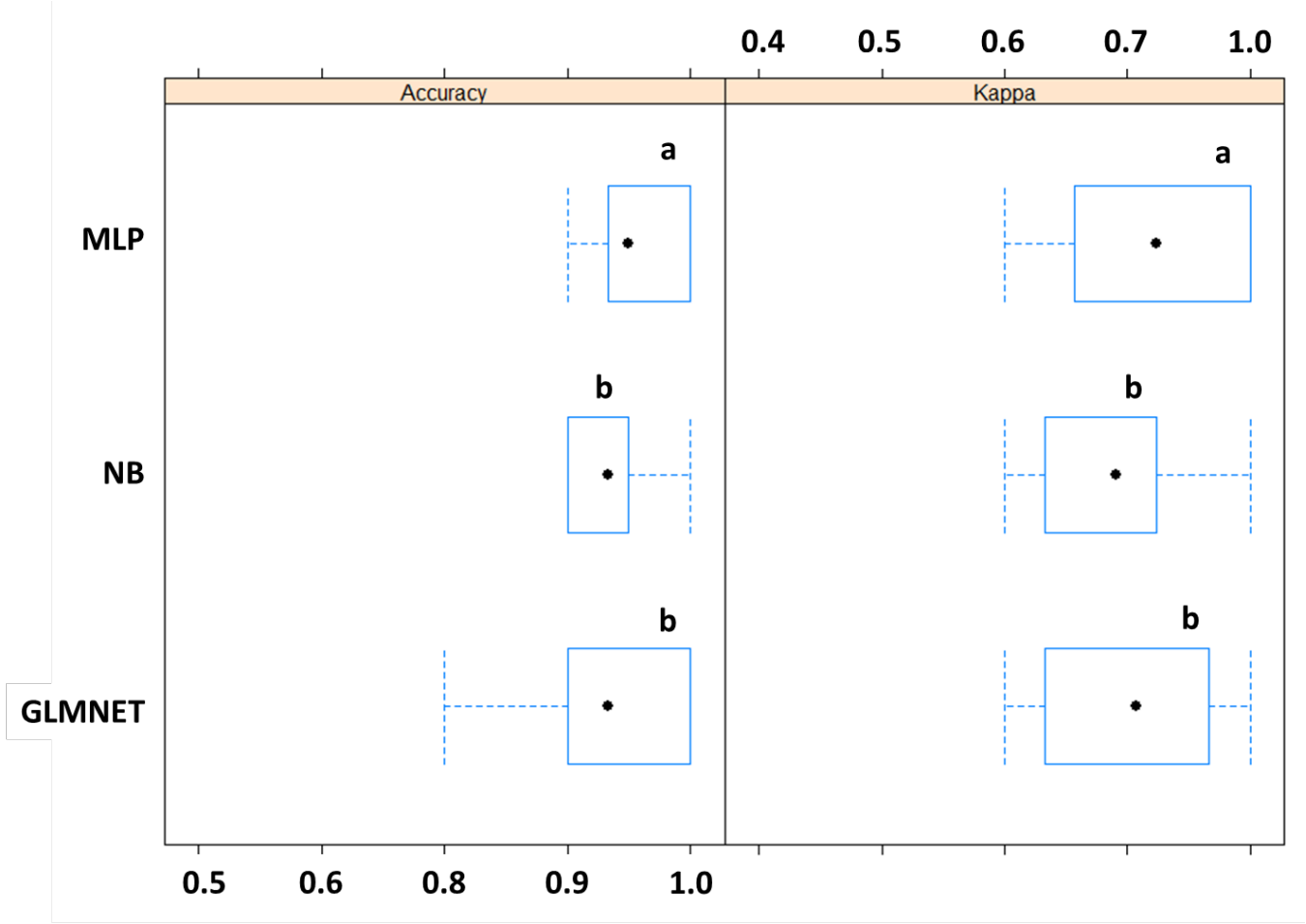


Figure S2.3. Comparative account of different three algorithms used for pectin classification using FT-IR spectra (Model II): multilayer perceptron (MLP, artificial neural network), naïve bayes classifier (NB) and generalized linear model elastic-net (GLMNET). Differences between models were calculated via their resampling distributions (number of resamples = 10). [a,b] Statistically significant differences between algorithms.

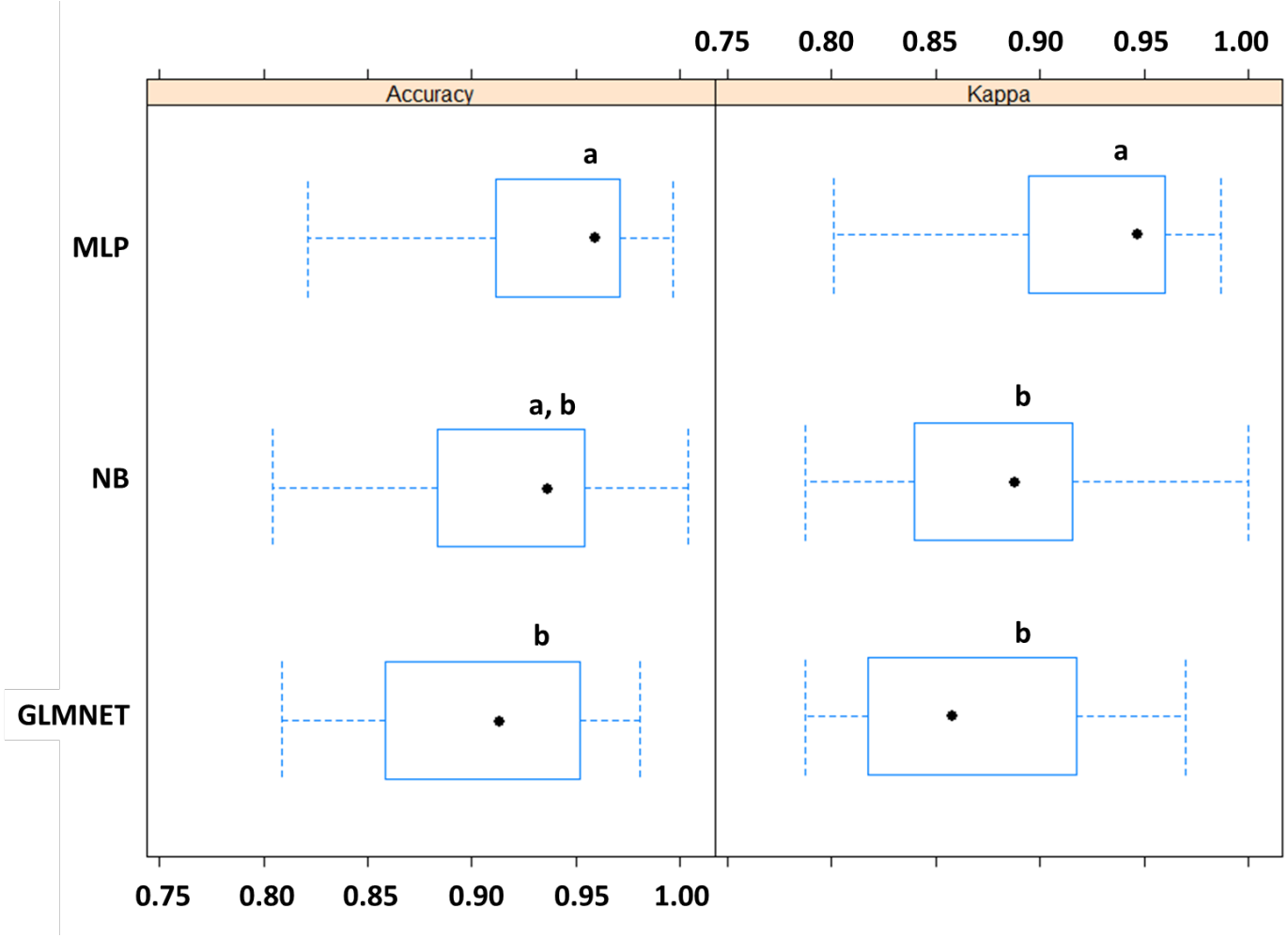


Figure S2.4. Comparative account of different three algorithms used for GC-MS spectra classification (Model III): multilayer perceptron (MLP, artificial neural network), naïve bayes classifier (NB) and generalized linear model elastic-net (GLMNET). Differences between models were calculated via their resampling distributions (number of resamples = 10). [a,b] Statistically significant differences between algorithms.

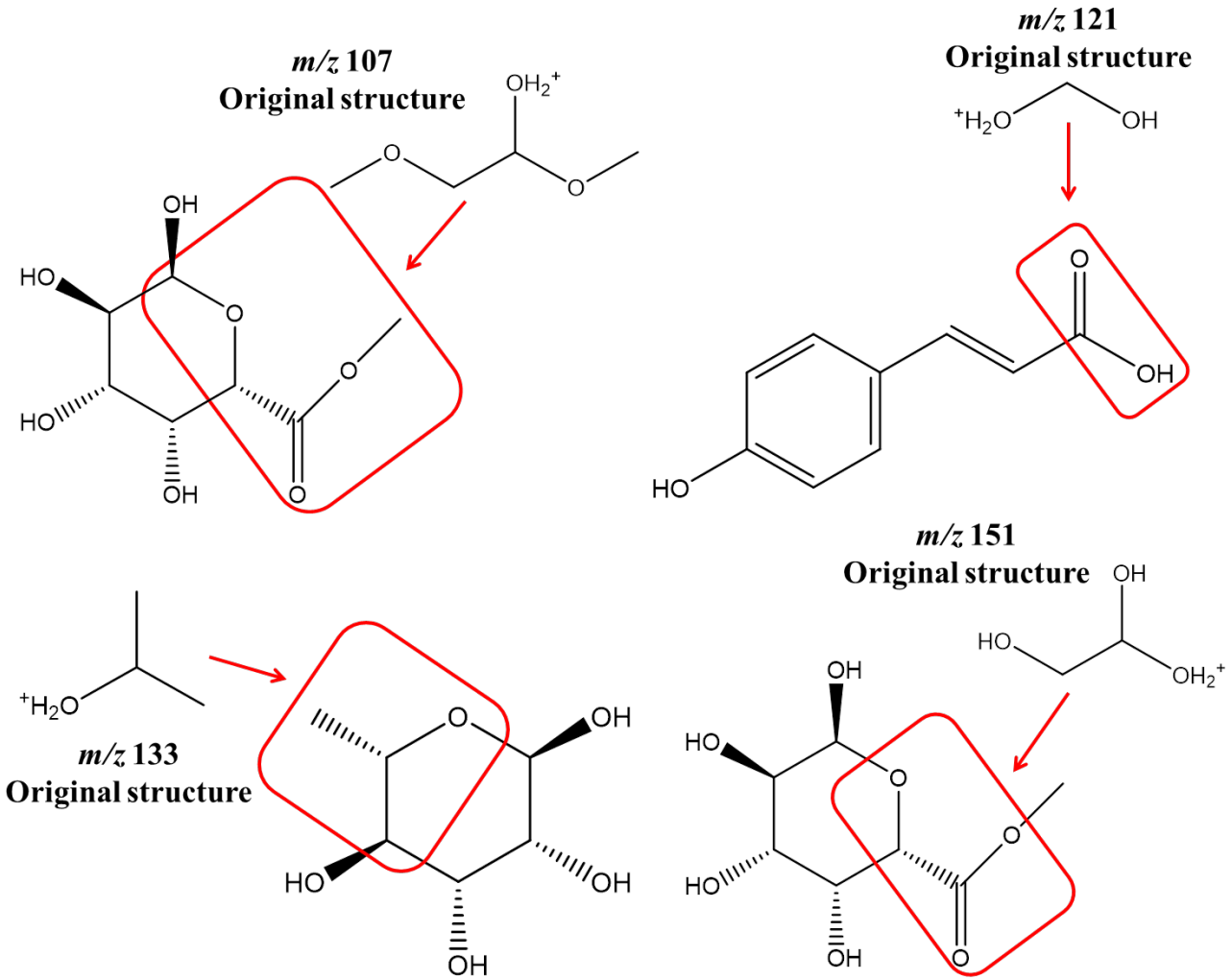


Figure S2.5. Chemical origin of relevant *m/z* fragments present in GC-MS spectra of studied compounds. Only original fragment structures calculated in Table 4 are considered.

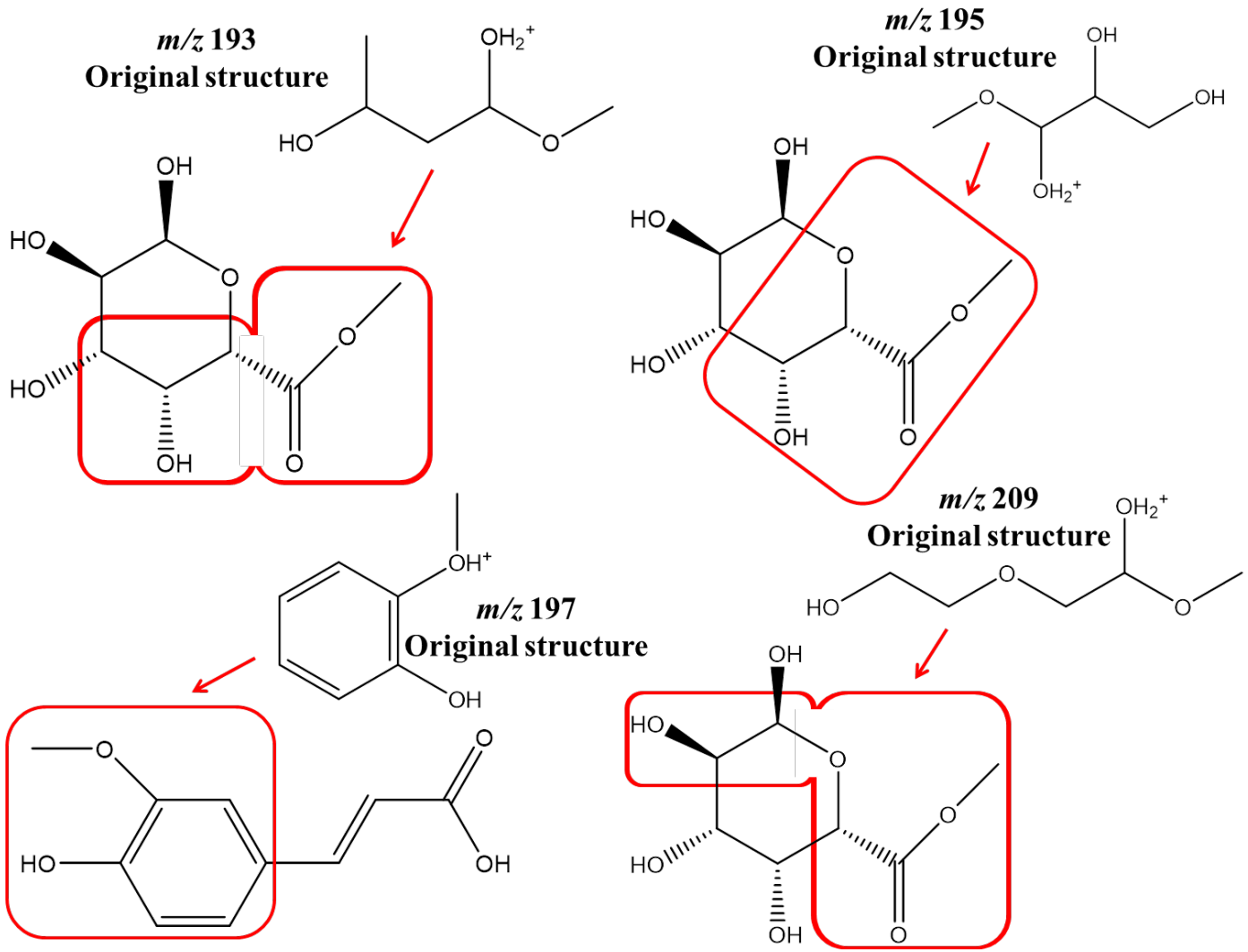


Figure S2.5. Cont.

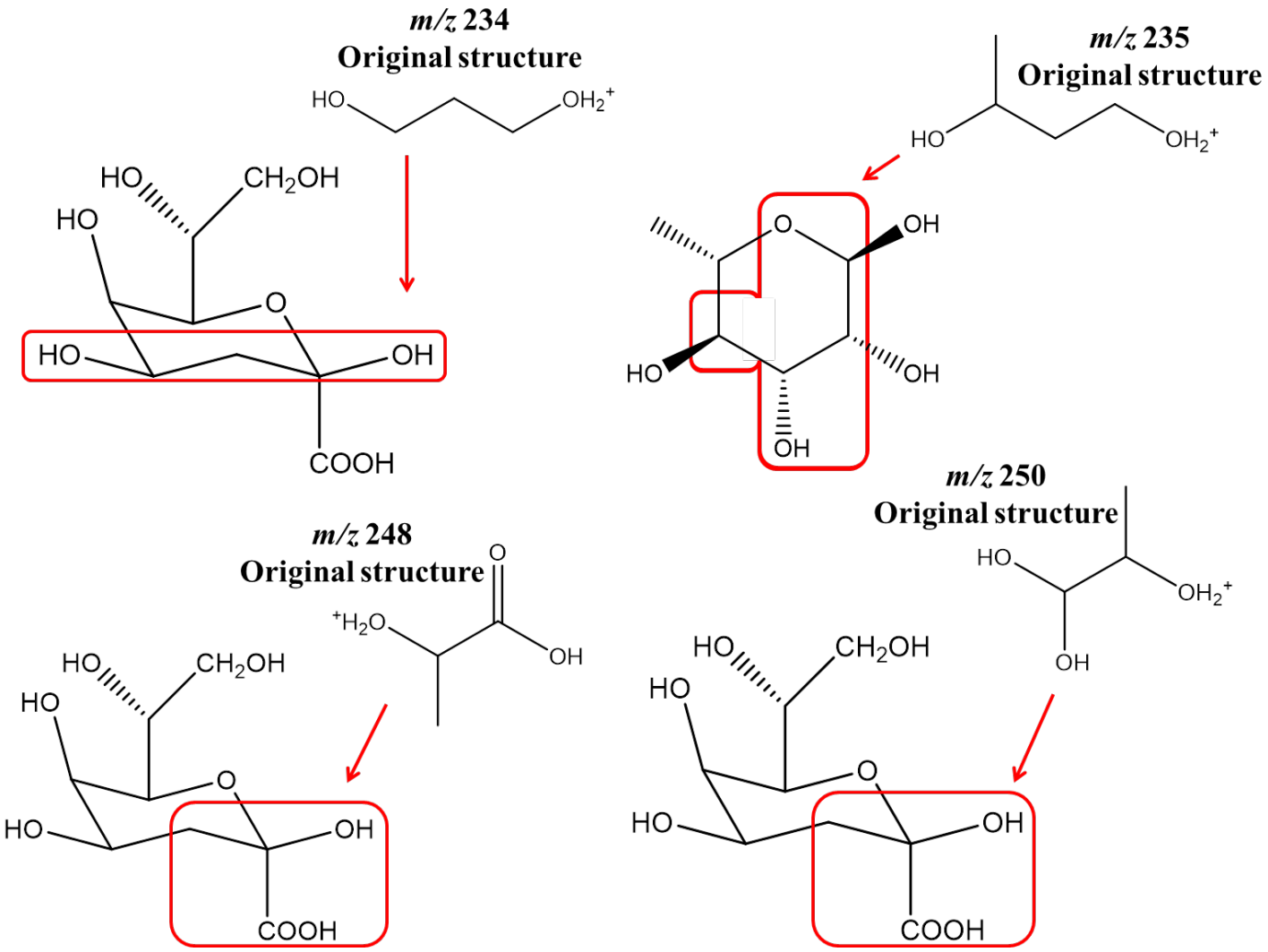


Figure S2.5. Cont.

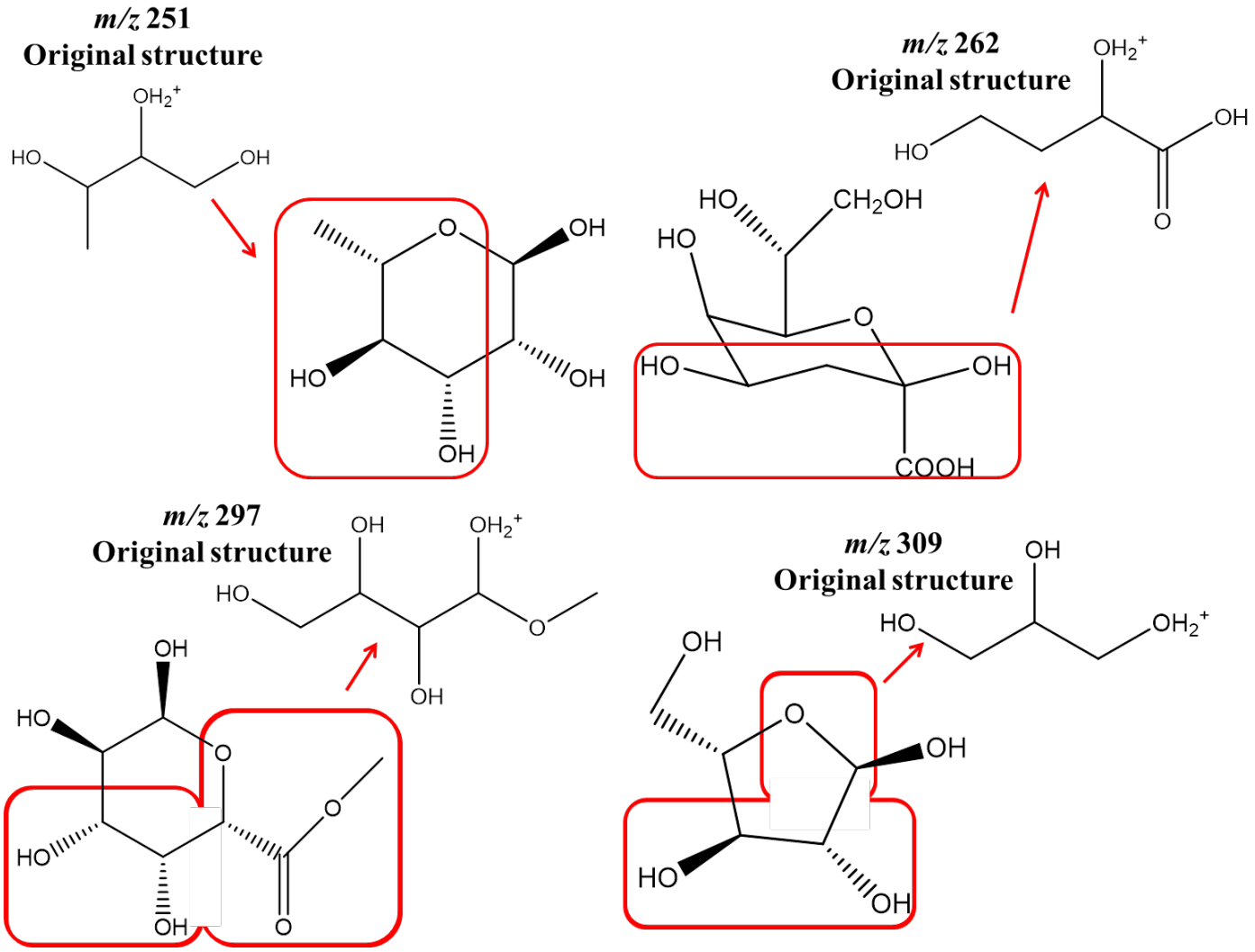


Figure S2.5. Cont.

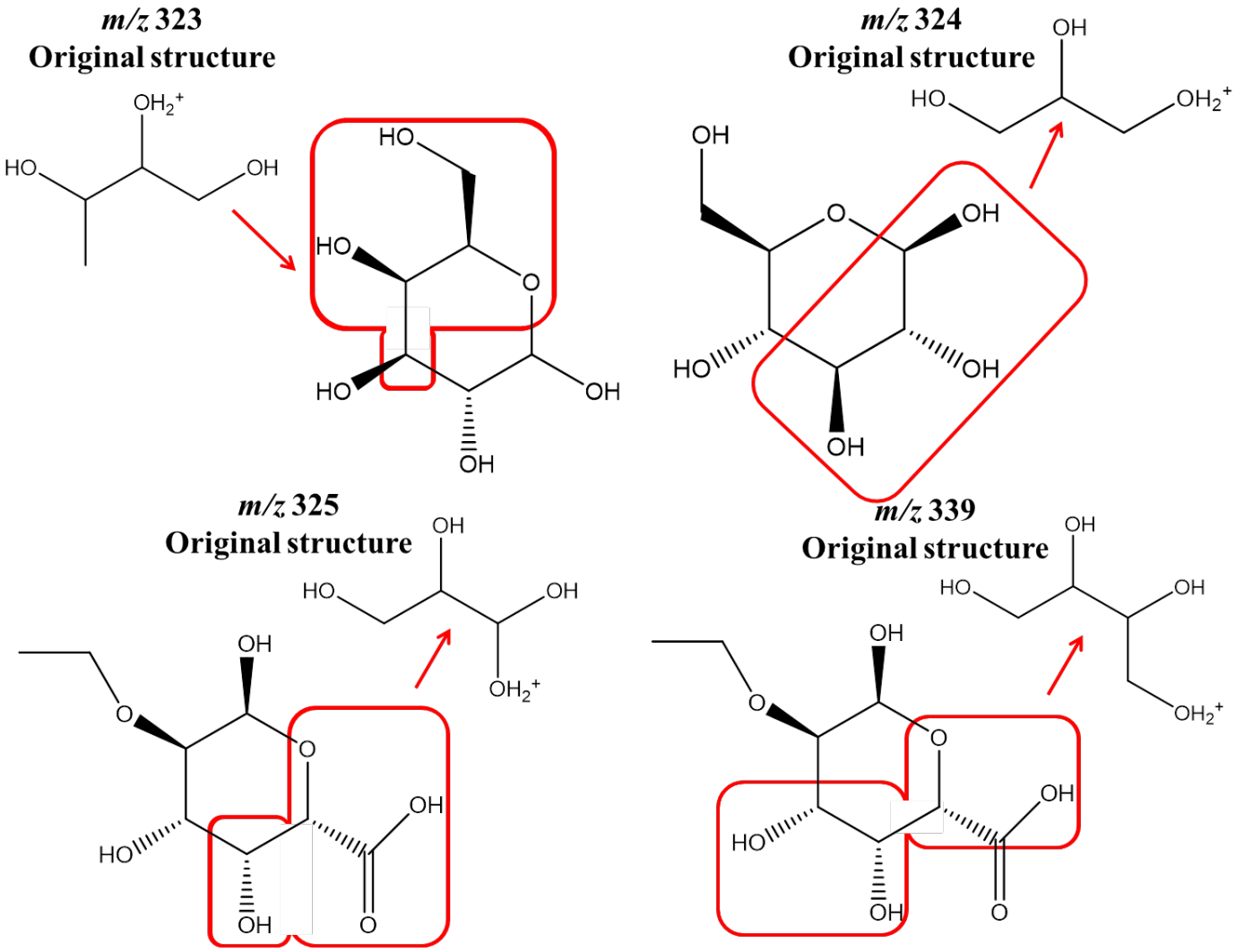


Figure S2.5. Cont.

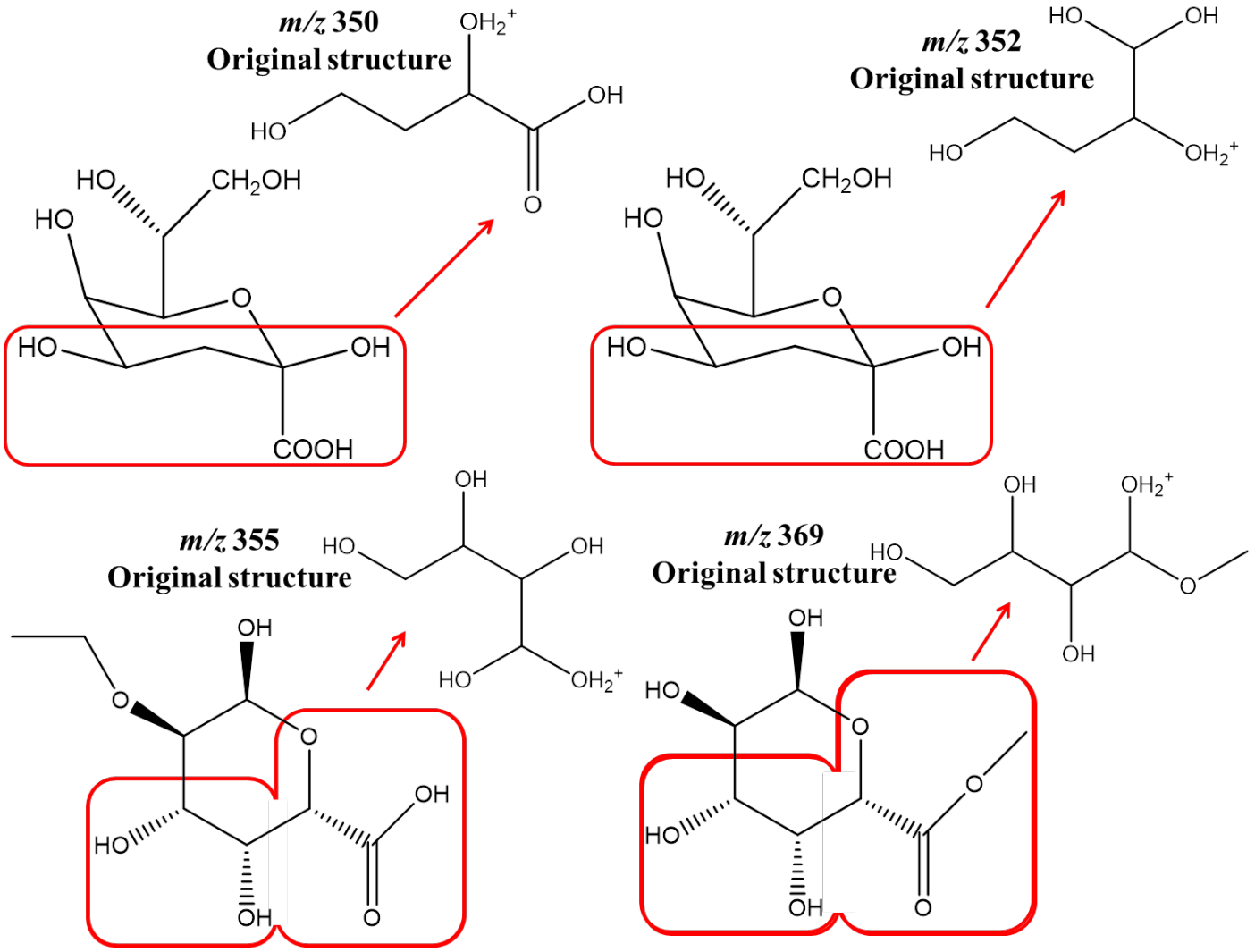


Figure S2.5. Cont.

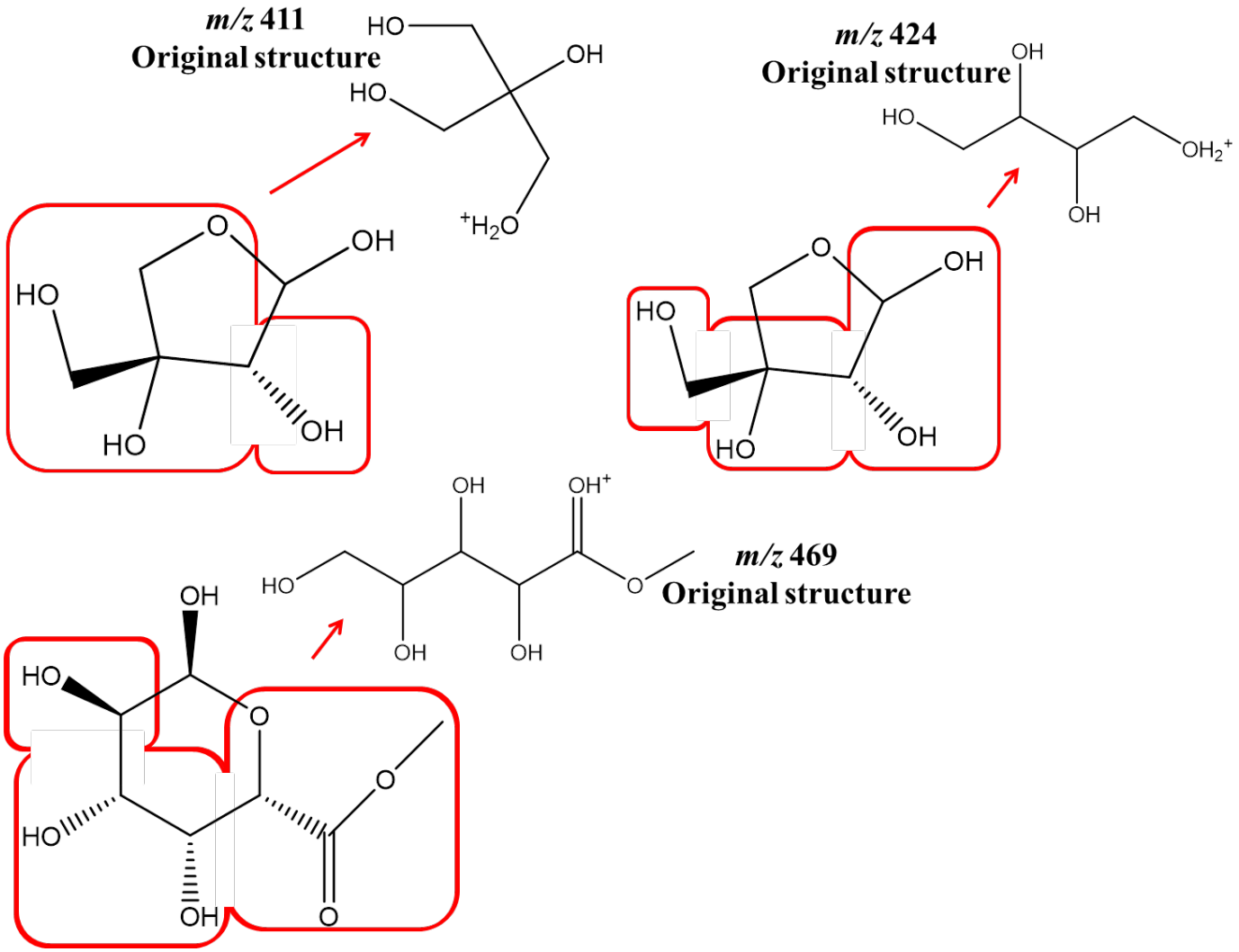


Figure S2.5. Cont.

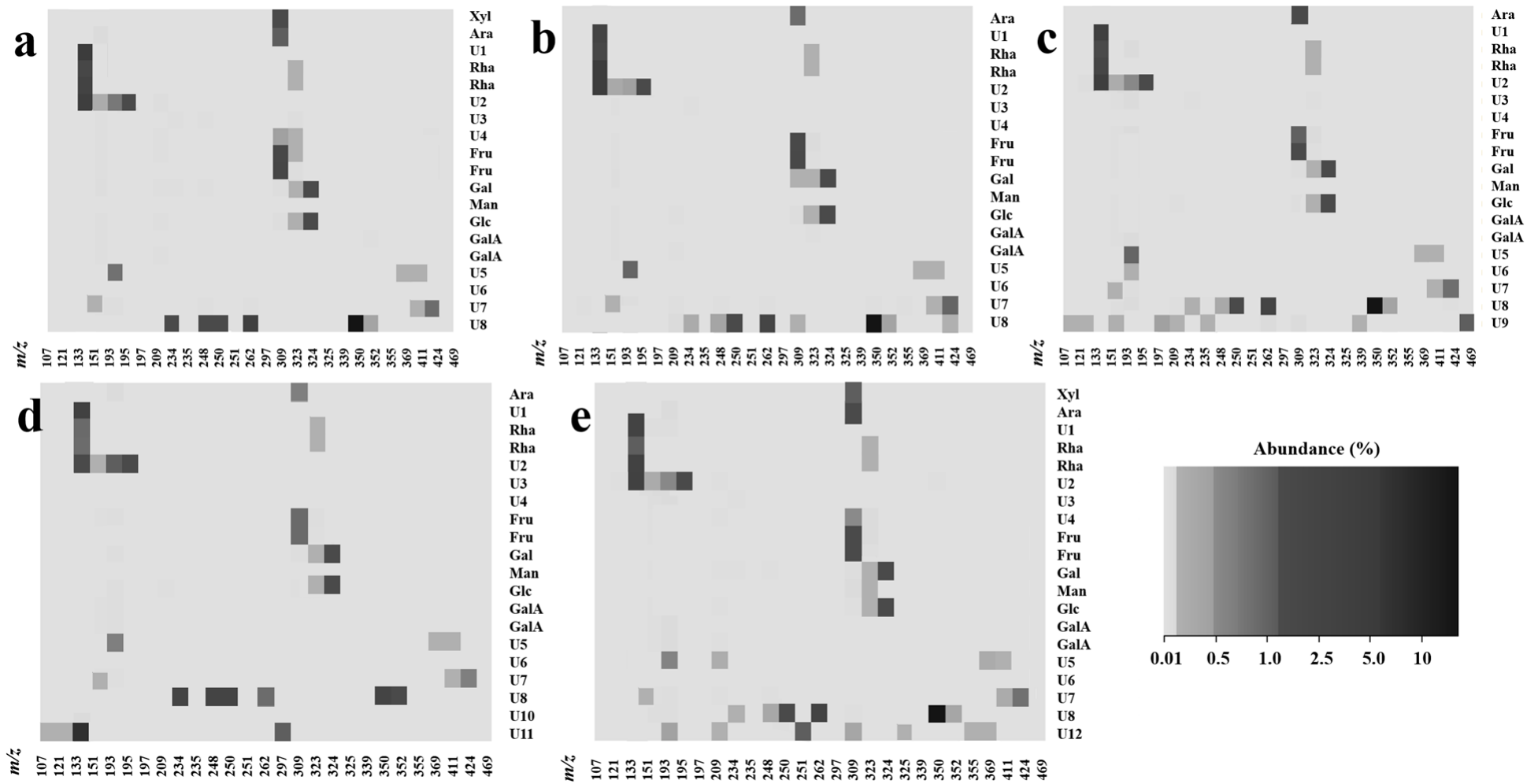


Figure S2.6. Heatmaps showing the abundances of relevant *m/z* ions present in GC-MS spectra of identified and unidentified monomers present in pectin extracted with: **a)** Celluclast, **b)** citrate, **c)** nitric acid, **d)** ultrasounds, **e)** combination of Celluclast and ultrasounds.

Annex I.III:

Supplementary material of Article III

11.3. Annex I.III: Supplementary material of Article III

Table S3.1. Most influential *m/z* ions in artichoke pectic oligosaccharides (POS) classification according to the enzyme selected by application of multilayer perceptron (MLP). Importance coefficients were determined by the sum of the product of raw input-hidden and hidden-output connection weights.

Known pectic sugars*		**Glucanex®200G**		**Cellulase from *A. niger***		**Pentopan®Mono-BG**		**Pectinex®Ultra-Olio**	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
207	100.0	**542**	100.0	**441**	100.0	**237**	100.0	**424**	100.0
309	74.2	**484**	94.9	**469**	88.5	**425**	82.8	**340**	83.9
502	67.7	**90**	89.1	**471**	88.0	**484**	78.6	**353**	83.8
385	62.6	**89**	83.6	**397**	85.4	**163**	77.8	**283**	78.3
513	60.4	**63**	82.9	**338**	79.8	**499**	67.3	**366**	74.7
281	58.9	**354**	74.0	**384**	65.8	**150**	65.6	**296**	73.7
195	57.0	**396**	68.1	**368**	63.9	**485**	63.3	**252**	72.3
456	56.5	**93**	65.5	**396**	63.6	**515**	60.4	**339**	67.8
338	55.5	**340**	64.2	**502**	62.9	**399**	59.0	**443**	67.7
339	53.9	**123**	63.3	**544**	62.3	**516**	58.4	**500**	67.0
426	53.5	**280**	60.2	**105**	58.2	**235**	58.3	**337**	64.9
180	53.1	**370**	58.4	**444**	57.8	**178**	54.4	**383**	63.4
528	49.5	**483**	57.6	**381**	53.3	**133**	53.9	**238**	62.5
472	47.0	**149**	55.0	**398**	52.9	**119**	53.4	**499**	60.7
471	45.2	**135**	53.2	**482**	52.5	**428**	49.8	**267**	60.4
544	44.4	**165**	52.0	**457**	51.6	**514**	49.1	**351**	60.3
310	44.3	**92**	51.7	**382**	51.3	**106**	43.2	**543**	56.9
225	43.2	**369**	51.4	**79**	50.2	**179**	43.0	**223**	54.9
546	42.4	**266**	50.6	**428**	49.7	**529**	40.5	**209**	53.4
411	41.5	**323**	50.5	**468**	49.6	**513**	40.2	**309**	53.1
382	41.0	**177**	47.9	**163**	48.6	**180**	39.2	**279**	50.9
323	39.8	**251**	46.9	**458**	46.4	**488**	39.2	**79**	48.6
527	38.5	**134**	45.5	**440**	46.3	**487**	39.0	**181**	48.3
414	32.1	**472**	45.5	**253**	46.3	**409**	38.6	**222**	47.2
428	31.5	**532**	44.1	**87**	43.6	**543**	37.3	**367**	47.1
439	29.8	**208**	44.0	**238**	42.4	**266**	37.3	**324**	45.8
498	29.5	**454**	40.5	**515**	42.0	**398**	33.8	**385**	44.5
177	28.3	**121**	38.5	**193**	38.5	**253**	33.3	**484**	43.1
341	28.1	**531**	34.8	**194**	36.7	**442**	31.1	**192**	40.2
311	27.6	**192**	33.4	**91**	35.0	**207**	29.3	**511**	38.7

*Known pectic sugars include galacturonic acid, arabinose, rhamnose, xylose, galactose, di- and trigalacturonic acid.

Table S3.2. Most influential *m/z* ions in artichoke pectic oligosaccharides (POS) classification according to the enzyme selected by application of random forest (RF). Importance coefficients were determined by the permutation of the out of-bag-error.

Known pectic sugars*		**Glucanex®200G**		**Cellulase from *A. niger***		**Pentopan®Mono-BG**		**Pectinex®Ultra-Olio**	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
546	100.0	**513**	100.0	**546**	100.0	**546**	100.0	**546**	100.0
472	50.0	**528**	93.5	**513**	71.2	**513**	80.0	**528**	81.3
513	47.1	**502**	79.8	**502**	61.9	**472**	76.3	**513**	81.3
502	41.5	**472**	77.1	**528**	59.5	**528**	72.9	**502**	78.5
530	39.2	**486**	75.8	**530**	56.1	**502**	71.3	**530**	76.8
458	33.9	**544**	71.0	**472**	56.0	**486**	66.4	**458**	71.9
501	29.6	**458**	70.0	**532**	53.5	**458**	62.4	**472**	71.4
528	29.6	**530**	64.8	**486**	51.5	**544**	57.7	**486**	65.1
399	27.4	**546**	63.0	**501**	50.7	**532**	55.4	**500**	57.4
544	26.7	**545**	45.4	**544**	50.6	**484**	49.6	**544**	51.0
483	24.2	**353**	45.3	**399**	48.4	**498**	47.6	**532**	49.4
532	23.7	**501**	43.3	**337**	47.9	**501**	47.3	**427**	49.4
485	23.3	**500**	42.4	**484**	42.1	**500**	44.7	**484**	47.6
500	22.2	**482**	41.8	**498**	40.3	**382**	42.9	**428**	46.4
382	22.0	**382**	41.4	**500**	39.5	**545**	41.3	**501**	45.6
337	20.7	**471**	39.8	**485**	37.8	**485**	40.4	**469**	44.6
527	19.5	**442**	38.7	**499**	35.7	**353**	39.3	**498**	43.4
529	19.0	**444**	38.7	**527**	33.8	**444**	39.2	**545**	43.3
339	18.4	**485**	37.1	**483**	32.9	**483**	39.1	**483**	43.0
426	18.1	**354**	36.4	**353**	30.7	**499**	38.6	**444**	42.4
486	17.9	**498**	36.0	**458**	28.1	**383**	36.7	**382**	40.1
338	17.5	**484**	32.3	**469**	27.9	**482**	35.7	**397**	39.3
383	17.5	**499**	31.6	**382**	27.8	**337**	33.6	**531**	37.8
498	17.2	**414**	30.8	**428**	27.3	**426**	32.9	**485**	36.8
499	16.6	**457**	30.3	**354**	25.7	**530**	31.7	**456**	34.4
442	16.4	**428**	29.9	**482**	24.1	**456**	31.2	**399**	33.1
428	15.6	**470**	29.1	**339**	23.8	**354**	31.1	**337**	32.2
398	15.6	**456**	29.1	**529**	23.0	**310**	31.0	**543**	31.7
484	15.5	**337**	28.3	**545**	22.9	**471**	30.9	**515**	30.6
353	14.6	**294**	27.2	**398**	22.7	**414**	28.7	**542**	29.7

*Known pectic sugars include galacturonic acid, arabinose, rhamnose, xylose, galactose, di- and trigalacturonic acid.

Table S3.3. Most influential *m/z* ions in artichoke pectic oligosaccharides (POS) classification according to the enzyme selected by application of boosted logistic regression (BLR). Importance coefficients were determined by calculating the area under the ROC (Receiver Operating Characteristic) curve.

Known pectic sugars*		**Glucanex®200G**		**Cellulase from *A. niger***		**Pentopan®Mono-BG**		**Pectinex®Ultra-Olio**	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
486	100.0	**337**	100.0	**353**	100.0	**337**	100	**546**	100.0
500	100.0	**353**	100.0	**354**	100.0	**353**	100	**502**	99.7
528	100.0	**484**	100.0	**414**	100.0	**484**	100	**399**	96.3
531	100.0	**486**	100.0	**456**	100.0	**486**	100	**382**	95.9
545	100.0	**498**	100.0	**458**	100.0	**498**	100	**513**	95.6
502	99.7	**502**	100.0	**471**	100.0	**500**	100	**398**	95.3
444	99.5	**528**	100.0	**472**	100.0	**502**	100	**530**	94.1
485	98.7	**382**	99.8	**482**	100.0	**528**	100	**544**	93.7
427	98.3	**544**	98.4	**486**	100.0	**531**	100	**338**	92.8
382	98.0	**444**	98.1	**499**	100.0	**545**	100	**428**	92.8
483	97.9	**546**	97.7	**528**	100.0	**382**	99.8	**337**	92.2
546	97.7	**398**	97.5	**502**	99.7	**444**	99.5	**501**	91.8
544	97.1	**545**	97.4	**485**	99.6	**485**	98.7	**486**	89.0
501	95.8	**428**	96.7	**546**	97.7	**544**	98.4	**472**	88.9
513	95.6	**513**	95.6	**511**	97.0	**427**	98.3	**457**	87.5
398	95.3	**501**	95.1	**382**	95.9	**483**	97.9	**339**	87.0
530	94.7	**530**	94.1	**545**	95.6	**398**	97.5	**528**	87.0
458	93.2	**338**	92.8	**513**	95.6	**428**	96.7	**458**	86.7
353	92.9	**457**	92.4	**470**	95.4	**501**	95.8	**353**	84.7
338	92.8	**471**	92.4	**398**	95.3	**546**	95.5	**426**	84.3
428	92.8	**426**	91.9	**444**	94.4	**530**	94.7	**383**	80.0
516	91.7	**500**	91.7	**530**	94.1	**513**	94.4	**444**	79.8
472	90.9	**499**	89.8	**500**	93.7	**458**	93.2	**532**	78.3
470	90.9	**412**	89.5	**544**	93.7	**457**	92.4	**469**	75.7
397	90.6	**472**	88.9	**338**	92.8	**471**	92.4	**471**	74.9
457	87.5	**458**	86.7	**428**	92.8	**426**	91.9	**545**	74.9
482	84.7	**532**	85.8	**501**	91.8	**516**	91.7	**252**	73.0
309	84.5	**339**	81.1	**412**	91.7	**472**	90.9	**411**	70.5
426	84.3	**383**	80.0	**442**	90.0	**470**	90.9	**296**	66.1
468	83.3	**531**	79.7	**457**	87.5	**397**	90.6	**531**	65.9

*Known pectic sugars include galacturonic acid, arabinose, rhamnose, xylose, galactose, di- and trigalacturonic acid.

Table S3.4. Possible chemical structures of other relevant *m/z* ions given by machine learning algorithms, determined by competitive fragment modelling (CFM-ID). These ions correspond to TMSO fragments from pectic di- or trisaccharides.

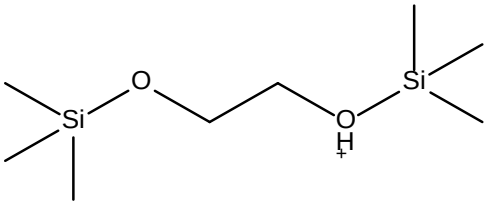
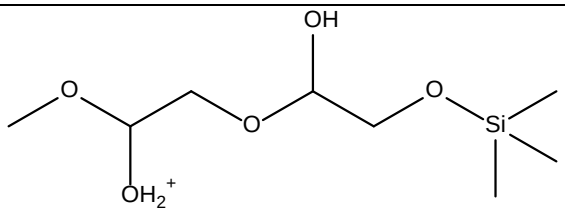
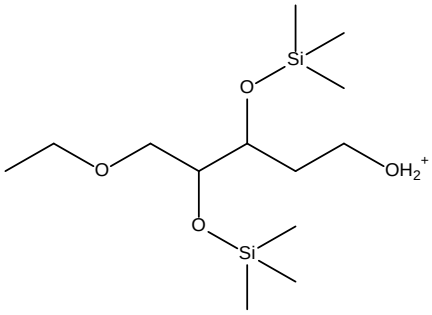
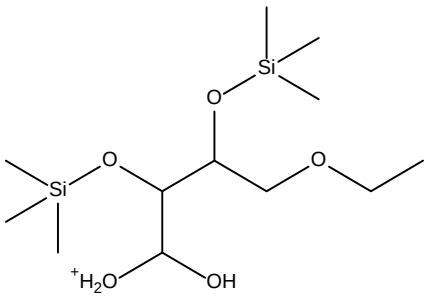
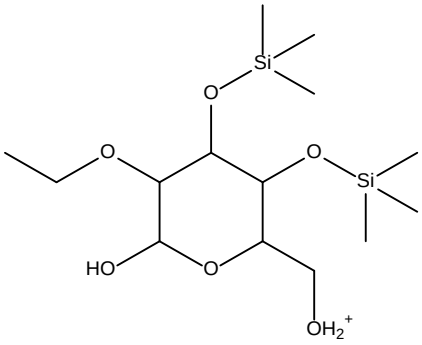
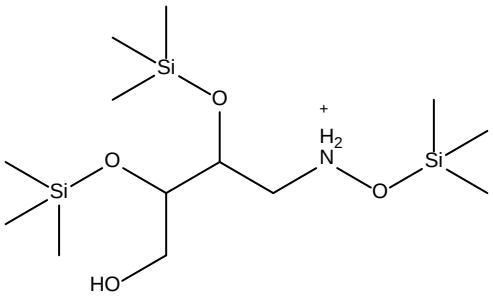
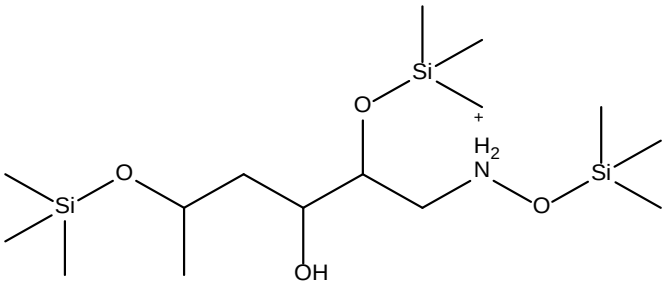
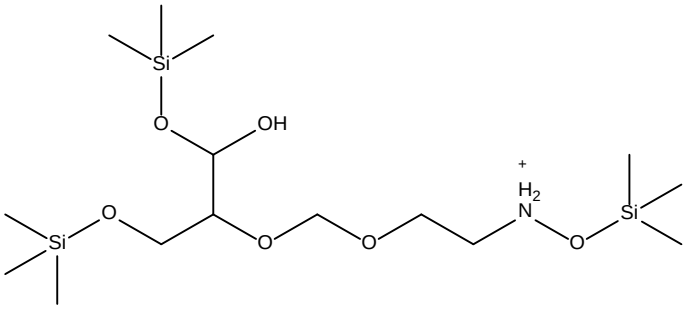
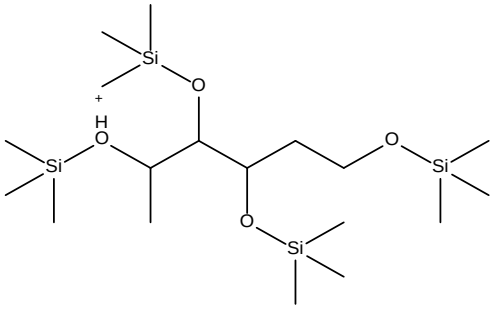
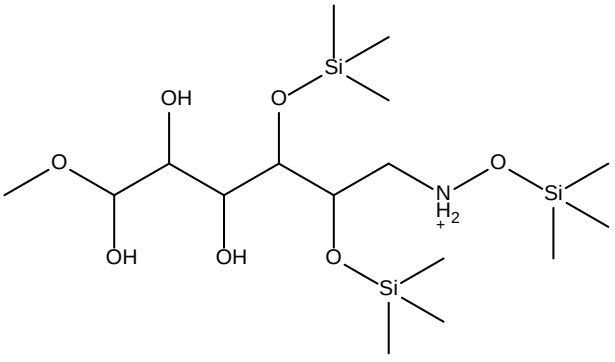
m/z	Structure	*m/z*	Structure
207		**225**	
309		**311**	
353		**354**	
382		**414**	
439		**444**	

Table S3.4. Cont.

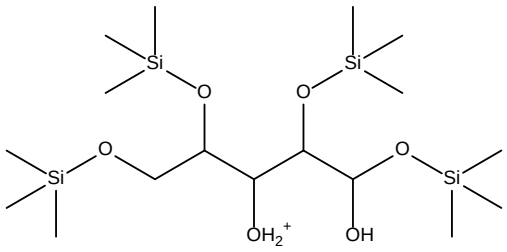
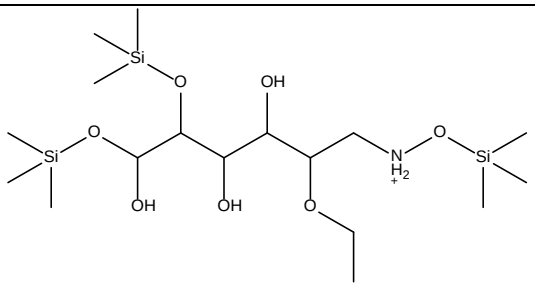
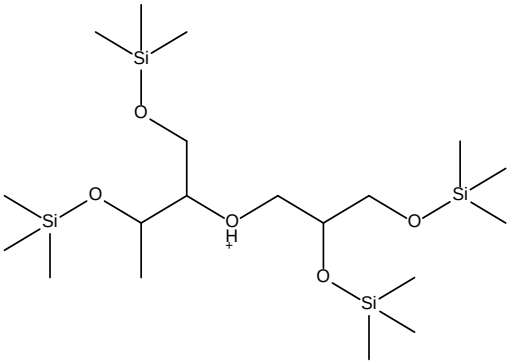
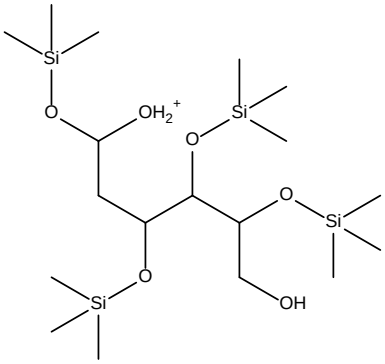
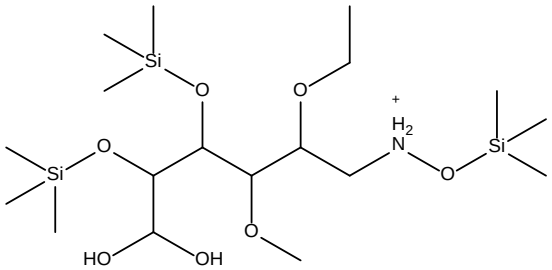
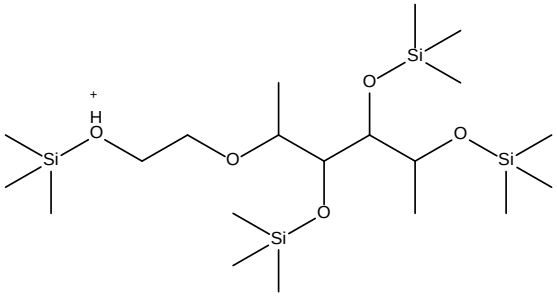
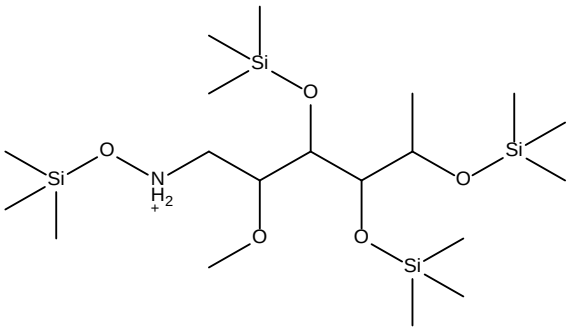
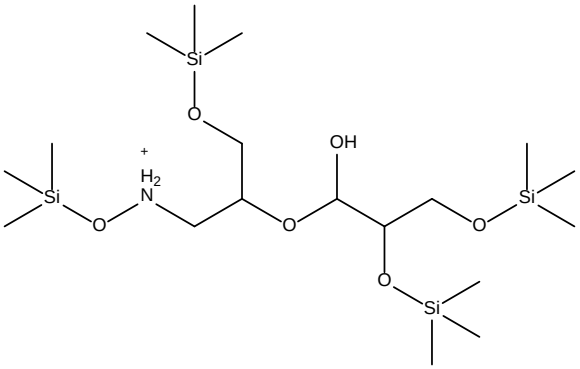
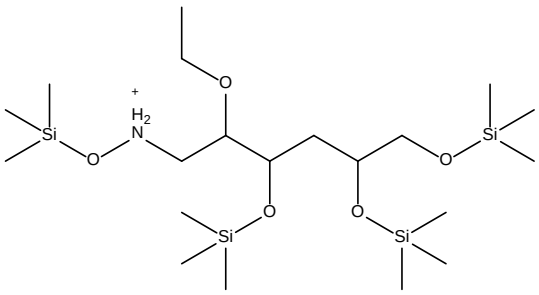
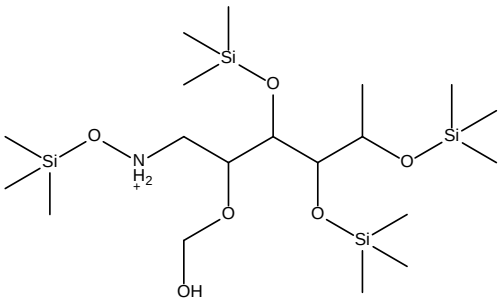
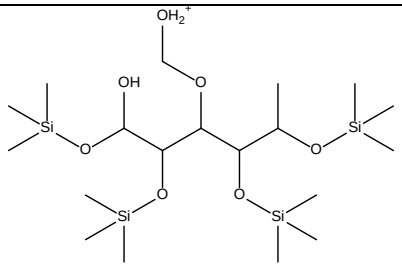
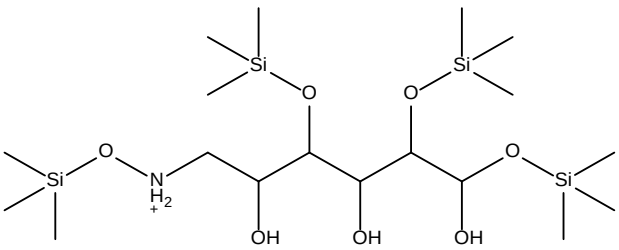
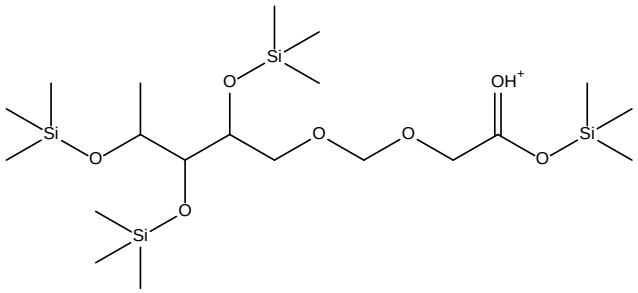
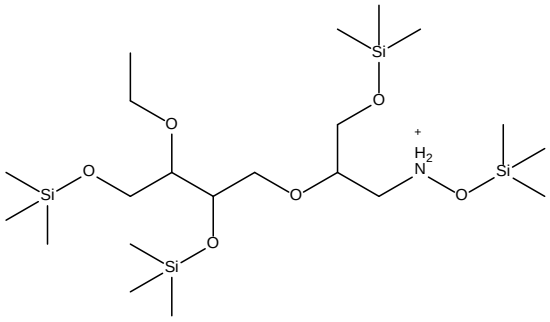
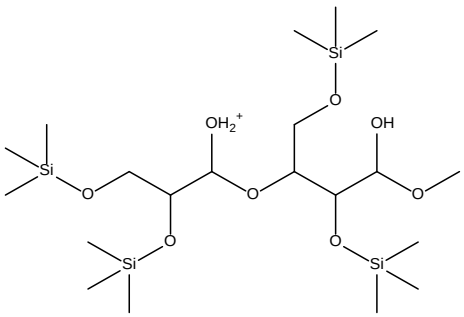
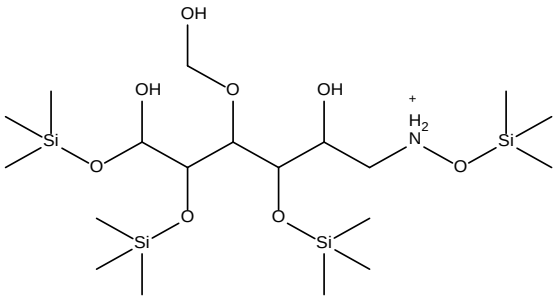
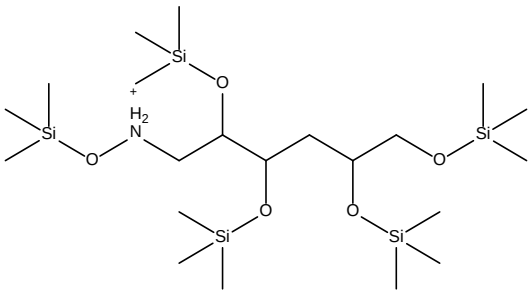
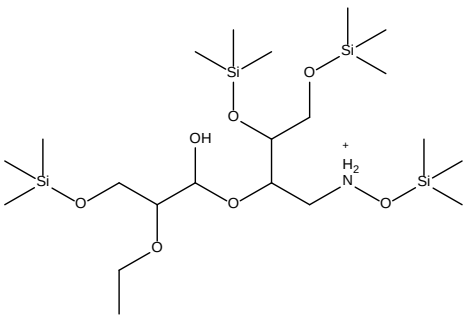
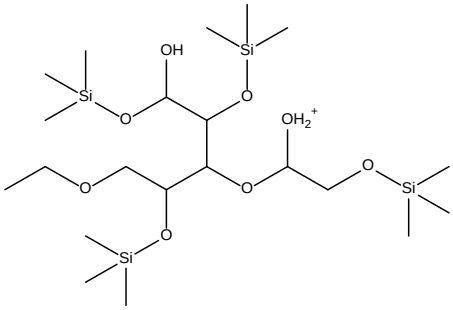
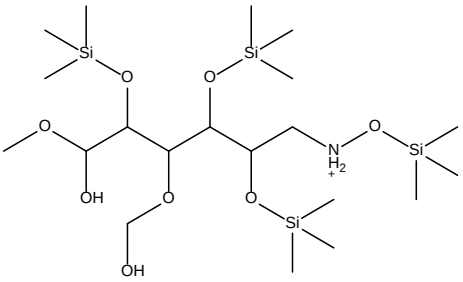
m/z	Structure	*m/z*	Structure
457		**458**	
469		**471**	
472		**483**	
484		**486**	
498		**500**	

Table S3.4. Cont.

m/z	Structure	*m/z*	Structure
501		**502**	
513		**528**	
531		**532**	
542		**544**	
545		**546**	

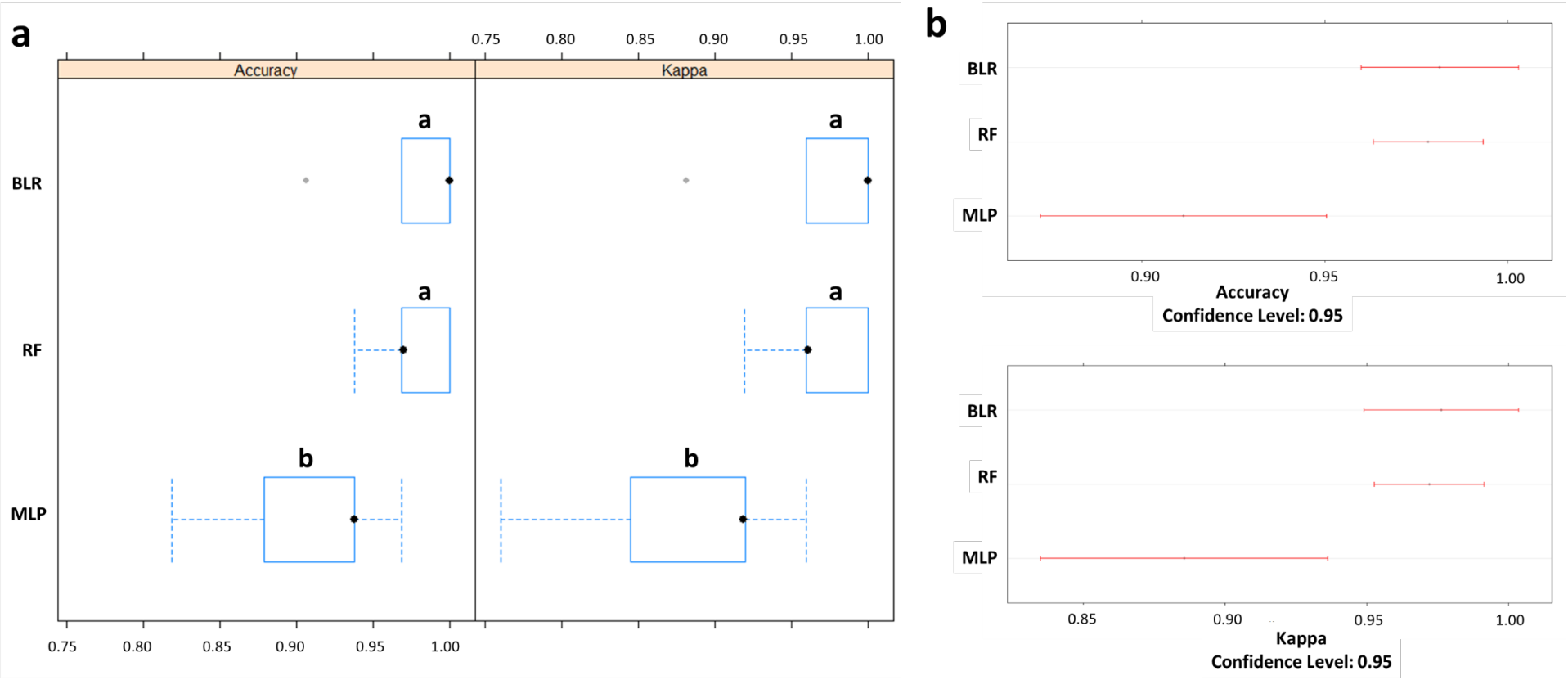


Figure S3.1. Comparative account of the three machine learning algorithms boosted logistic regression (BLR), random forest (RF) and multilayer perceptron (MLP): **a)** boxplots of accuracy and kappa values for each models, **b)** accuracy and kappa values obtained at a confidence level of 0.95. Differences between models were calculated via their resampling distributions (number of resamples = 10).

[a,b] Statistically significant differences between algorithms.

Annex I.IV:

Supplementary material of Article IV

11.4. Annex I.IV: Supplementary material of Article IV

Table S4.1. Number of GC-MS spectra (n = 212) of pectic sugars and oligosaccharides obtained from GC-EI-MS analysis of hydrolysates from pectin and polygalacturonic acid included in machine learning study. GalA: galacturonic acid, Di-GalA: digalacturonic acid, Tri-GalA: trigalacturonic acid, POS: pectic-oligosaccharides.

Commercial enzyme	Substrate	Number of hydroslysates	Number of GC-MS spectra					
			Known pectic monosaccharides	Di-GalA	Unknown POS-DP2	Tri-GalA	Unknown POS-DP3	Total
Glucanex	Citrus	2	-	4 (2 per hydrolysate)	8 (4 per hydrolysate)	4 (2 per hydrolysate)	4 (2 per hydrolysate)	8 Di-, Tri-GalA 12 unknown POS
	Apple	2	-	4 (2 per hydrolysate)	12 (6 per hydrolysate)	4 (2 per hydrolysate)	6 (3 per hydrolysate)	8 Di-, Tri-GalA 18 unknown POS
	Polygalacturonic acid	2	-	4 (2 per hydrolysate)	24 (12 per hydrolysate)	4 (2 per hydrolysate)	16 (8 per hydrolysate)	8 Di-, Tri-GalA 40 unknown POS
Viscozyme	Citrus	2	-	4 (2 per hydrolysate)	18 (9 per hydrolysate)	4 (2 per hydrolysate)	18 (9 per hydrolysate)	8 Di-, Tri-GalA 36 unknown POS
	Apple	2	-	4 (2 per hydrolysate)	14 (7 per hydrolysate)	4 (2 per hydrolysate)	8 (4 per hydrolysate)	8 Di-, Tri-GalA 22 unknown POS
	Polygalacturonic acid	2	-	4 (2 per hydrolysate)	18 (9 per hydrolysate)	4 (2 per hydrolysate)	4 (2 per hydrolysate)	8 Di-, Tri-GalA 22 unknown POS
-	Standard	-	10[1]	2[2]	-	2[3]	-	14 standards

[1] Known monosaccharides including galacturonic acid, arabinose, rhamnose, xylose and galactose. [2] Digalacturonic acid standards. [3]Trigalacturonic acid standards.

Table S4.2. Most influential *m/z* ions in artichoke pectic oligosaccharides (POS) classification according to the substrate selected by application of random forest (RF). Importance coefficients were determined by the permutation of the out of-bag-error. PGA: polygalacturonic acid.

Known pectic sugars		**Apple pectin**		**Citrus pectin**		**PGA**	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
747	99.5	**528**	90.1	**603**	93.8	**617**	100.0
617	95.6	**629**	89.9	**370**	92.0	**747**	99.2
603	95.4	**428**	89.0	**484**	89.8	**603**	98.8
629	90.4	**616**	88.6	**706**	87.9	**746**	98.5
646	90.0	**720**	88.1	**121**	87.9	**121**	98.2
675	87.4	**706**	87.5	**424**	85.7	**370**	97.9
223	87.4	**558**	86.9	**615**	85.1	**514**	94.7
746	87.1	**734**	86.0	**658**	85.0	**691**	90.4
571	83.9	**615**	85.6	**733**	83.4	**720**	87.2
699	83.0	**658**	84.8	**630**	83.4	**675**	85.6
179	82.3	**630**	83.9	**719**	83.2	**706**	84.8
558	82.3	**515**	83.1	**223**	82.4	**530**	84.5
719	82.1	**689**	82.2	**704**	82.2	**482**	84.0
106	82.0	**546**	80.1	**235**	82.0	**571**	83.9
633	81.5	**107**	79.9	**501**	81.9	**616**	83.0
615	81.4	**414**	79.3	**703**	80.8	**528**	82.9
657	80.5	**747**	79.2	**252**	80.7	**715**	82.8
546	79.8	**413**	77.7	**616**	80.4	**527**	82.7
454	74.4	**715**	77.6	**556**	80.4	**719**	82.0
674	74.0	**603**	77.6	**514**	80.3	**689**	81.9
119	73.8	**482**	77.5	**515**	80.0	**515**	81.7
267	73.5	**703**	77.3	**747**	79.4	**629**	81.5
107	72.3	**646**	77.2	**179**	79.2	**704**	80.1
689	71.9	**121**	77.0	**657**	78.1	**657**	80.1
703	71.9	**235**	76.4	**746**	77.9	**513**	79.6
178	71.9	**223**	76.4	**571**	77.7	**558**	79.1
239	71.5	**733**	76.3	**458**	77.5	**107**	78.1
484	71.1	**604**	76.2	**617**	77.5	**674**	77.7
691	70.9	**514**	75.8	**558**	76.9	**457**	77.5
720	70.9	**530**	75.6	**428**	76.6	**703**	77.3

Table S4.3. Most influential *m/z* ions in artichoke pectic oligosaccharides (POS) classification according to the substrate selected by application of boosted logistic regression (BLR). Importance coefficients were determined by calculating the area under the ROC (Receiver Operating Characteristic) curve. PGA: polygalacturonic acid.

Known pectic sugars		**Apple pectin**		**Citrus pectin**		**PGA**	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
703	98.4	**616**	100.0	**616**	100.0	**616**	100.0
616	97.6	**703**	98.4	**703**	97.7	**703**	97.7
699	95.1	**658**	96.3	**658**	96.3	**658**	96.3
747	91.1	**630**	94.4	**699**	95.1	**630**	94.4
734	87.0	**734**	88.8	**630**	94.4	**734**	88.8
658	84.9	**715**	83.5	**747**	91.1	**733**	82.4
630	84.5	**720**	82.4	**734**	88.8	**428**	78.8
715	83.5	**428**	78.8	**657**	81.7	**613**	73.0
720	82.4	**370**	76.3	**179**	80.2	**647**	72.8
657	81.7	**647**	74.5	**428**	78.8	**720**	72.6
179	80.2	**613**	73.0	**629**	78.3	**603**	70.4
629	78.3	**604**	71.3	**546**	77.4	**442**	67.9
546	77.4	**674**	70.4	**571**	76.8	**121**	67.7
571	76.8	**731**	65.4	**133**	75.9	**514**	67.1
370	76.3	**603**	64.5	**239**	75.2	**424**	66.5
133	75.9	**164**	61.9	**603**	73.1	**617**	65.6
239	75.2	**324**	61.6	**613**	73.0	**370**	64.1
647	74.5	**646**	61.6	**647**	72.8	**706**	62.4
603	73.1	**426**	60.5	**720**	72.6	**164**	61.9
558	72.4	**691**	57.8	**558**	72.4	**715**	61.9
106	72.1	**556**	57.0	**106**	72.1	**324**	61.6
646	71.8	**617**	56.0	**646**	71.8	**646**	61.6
604	71.3	**675**	56.0	**675**	70.6	**426**	60.5
675	70.6	**689**	55.8	**470**	69.2	**731**	60.2
674	70.4	**134**	54.5	**633**	68.3	**604**	57.0
470	69.2	**706**	54.4	**223**	66.7	**556**	57.0
633	68.3	**657**	53.3	**719**	66.5	**615**	54.9
223	66.7	**747**	52.8	**267**	66.1	**134**	54.5
719	66.5	**558**	51.3	**617**	63.1	**484**	53.6
267	66.1	**515**	49.8	**164**	61.9	**657**	53.3

Table S4.4. Most influential *m/z* ions in artichoke pectic oligosaccharides (POS) classification according to the enzyme selected by application of random forest (RF). Importance coefficients were determined by the permutation of the out of-bag-error.

Known pectic sugars		**Glucanex**		**Viscozyme**	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
660	95.9	**660**	100.0	**660**	91.6
704	88.1	**309**	81.7	**309**	88.5
616	78.8	**310**	80.0	**604**	83.0
719	76.9	**442**	79.4	**442**	81.8
734	75.8	**123**	79.3	**135**	81.4
613	75.3	**603**	78.2	**513**	80.3
630	75.1	**604**	77.4	**238**	78.9
733	74.7	**733**	77.3	**689**	78.4
731	73.4	**513**	75.9	**528**	78.4
699	73.0	**546**	75.5	**516**	77.4
309	72.1	**674**	74.8	**235**	76.4
442	71.3	**235**	73.8	**704**	75.2
151	70.9	**749**	73.4	**121**	75.1
617	70.6	**528**	72.7	**501**	74.3
749	70.5	**294**	72.6	**123**	74.2
675	69.1	**704**	71.7	**630**	74.1
516	68.9	**458**	71.6	**252**	74.0
238	68.5	**252**	70.6	**719**	73.9
106	68.4	**732**	70.3	**236**	73.9
747	67.9	**514**	69.3	**749**	73.8
715	67.5	**511**	69.2	**546**	73.6
674	67.4	**558**	68.7	**617**	73.5
484	67.0	**715**	68.5	**502**	73.4
121	66.4	**502**	68.5	**675**	72.9
382	65.7	**734**	67.8	**107**	72.8
251	65.6	**238**	67.7	**454**	72.2
604	64.6	**516**	66.5	**658**	72.0
239	63.9	**279**	66.5	**731**	71.5
123	63.1	**544**	66.1	**310**	71.3
720	63.0	**719**	65.9	**747**	70.0

Table S4.5. Most influential *m/z* ions in artichoke pectic oligosaccharides (POS) classification according to the enzyme selected by application of boosted logistic regression (BLR). Importance coefficients were determined by calculating the area under the ROC (Receiver Operating Characteristic) curve.

Known pectic sugars		**Glucanex**		**Viscozyme**	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
660	100.0	**660**	100.0	**660**	94.7
734	97.8	**734**	97.8	**704**	91.2
704	96.2	**704**	96.2	**309**	91.1
674	92.5	**674**	92.5	**617**	86.2
719	91.7	**719**	91.7	**528**	85.4
603	90.7	**309**	91.1	**279**	84.7
617	90.1	**603**	90.7	**616**	83.7
616	89.8	**617**	90.1	**719**	82.9
458	86.5	**616**	89.8	**699**	82.3
746	85.0	**458**	86.5	**294**	81.3
706	84.8	**528**	85.4	**675**	81.0
733	84.3	**746**	85.0	**454**	79.3
715	84.2	**706**	84.8	**715**	78.5
613	84.0	**279**	84.7	**747**	78.2
720	84.0	**733**	84.3	**238**	78.1
699	82.3	**715**	84.2	**310**	78.1
731	82.2	**613**	84.0	**731**	77.0
675	81.6	**720**	84.0	**123**	76.6
442	81.3	**731**	82.2	**235**	76.5
747	80.2	**675**	81.6	**658**	76.4
454	79.3	**294**	81.3	**613**	76.4
238	78.1	**442**	81.3	**657**	75.8
658	76.4	**747**	80.2	**151**	75.7
310	75.9	**310**	78.1	**106**	75.4
657	75.8	**123**	76.6	**513**	71.9
151	75.7	**235**	76.5	**734**	71.5
106	75.4	**699**	74.2	**179**	70.6
647	72.5	**238**	73.6	**180**	69.6
558	72.4	**647**	72.5	**703**	69.4
630	72.4	**558**	72.4	**442**	69.0

Table S4.6. Sensitivity, specificity, true positives and negatives and balanced accuracy rates (%) for multilayer perceptron (MLP) classification of POS according to both sample and enzyme source. PGA: polygalacturonic acid.

Parameter	Glucanex-Apple-POS	Viscozyme-Apple-POS	Glucanex-Citrus-POS	Viscozyme-Citrus-POS	Glucanex-PGA-POS	Viscozyme-PGA-POS	Known pectic sugars
Sensitivity	85.7	100	100	90.9	73.3	100	100
Specificity	100	100	100	98.2	98.0	96.7	95.9
True positives	100	100	100	90.9	91.7	71.4	89.5
True negatives	98.3	100	100	98.2	92.6	100	100
Balanced Accuracy	92.9	100	100	94.6	85.7	98.4	98.0

Table S4.7. Most influential *m/z* ions in artichoke pectic oligosaccharides (POS) classification according to both substrate and enzyme selected by application of random forest (RF). Importance coefficients were determined by the permutation of the out of-bag-error.

Known pectic sugars		**Glucanex-Apple**		**Viscozyme-Apple**		**Glucanex-Citrus**	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
223	100.0	**223**	67.3	**399**	52.3	**399**	39.2
252	63.3	**646**	54.5	**720**	49.6	**646**	38.9
107	50.5	**399**	52.3	**646**	49.3	**720**	35.9
457	38.3	**691**	51.9	**530**	42.3	**223**	31.8
121	36.9	**720**	51.4	**457**	42.1	**121**	31.2
617	35.2	**252**	48.0	**715**	37.9	**107**	30.6
482	34.9	**689**	38.9	**107**	37.1	**715**	28.8
720	31.0	**719**	38.3	**252**	35.1	**747**	28.3
472	29.9	**279**	37.9	**631**	35.0	**691**	26.8
251	29.4	**556**	36.5	**747**	34.8	**327**	25.5
296	29.1	**482**	36.2	**571**	31.6	**704**	24.9
571	28.9	**715**	32.7	**556**	28.1	**631**	24.7
715	28.2	**734**	31.8	**734**	27.7	**209**	24.4
134	26.8	**617**	31.7	**704**	26.3	**657**	23.9
327	26.7	**107**	31.5	**706**	26.1	**617**	23.4
222	26.5	**530**	30.7	**617**	25.9	**134**	22.6
747	26.0	**121**	30.1	**657**	25.8	**603**	22.1
151	24.8	**704**	29.6	**482**	25.8	**279**	22.0
239	24.8	**657**	29.3	**123**	25.5	**236**	21.9
235	24.4	**747**	27.8	**134**	25.3	**613**	21.9
135	23.8	**457**	27.3	**719**	24.9	**733**	21.7
657	23.4	**134**	26.9	**691**	24.9	**746**	21.6
691	23.1	**630**	26.8	**222**	24.1	**630**	21.5
704	22.0	**733**	26.7	**632**	24.0	**734**	21.0
647	22.0	**613**	26.2	**733**	23.9	**427**	20.7
646	21.9	**706**	26.1	**542**	23.4	**458**	20.4
556	21.8	**310**	24.6	**689**	23.3	**706**	19.4
283	21.7	**282**	22.5	**427**	22.7	**660**	18.9
630	21.4	**660**	21.3	**428**	22.1	**135**	18.8
719	20.7	**603**	21.1	**658**	21.9	**647**	18.8

Table S4.7. Cont.

Viscozyme-Citrus		Glucanex-PGA		Viscozyme-PGA	
m/z	Importance	*m/z*	Importance	*m/z*	Importance
223	85.6	**223**	85.9	**223**	52.0
646	78.5	**399**	60.9	**646**	51.6
720	68.3	**646**	58.4	**720**	51.1
457	44.6	**720**	51.9	**399**	50.0
734	40.3	**252**	50.1	**252**	50.0
252	38.8	**279**	46.2	**279**	39.5
631	31.6	**631**	43.7	**657**	39.0
617	30.9	**482**	42.4	**719**	37.3
399	30.7	**107**	36.9	**613**	36.4
472	28.4	**134**	35.8	**733**	34.9
134	26.0	**121**	34.9	**630**	34.0
706	24.8	**310**	32.6	**704**	33.5
691	24.0	**734**	32.5	**747**	31.4
632	23.4	**632**	28.0	**617**	30.2
310	22.7	**296**	26.6	**691**	30.1
613	22.6	**719**	25.5	**530**	29.2
603	22.5	**617**	25.4	**689**	28.5
107	22.2	**715**	22.7	**121**	27.5
482	22.0	**514**	22.6	**472**	27.4
544	19.7	**704**	22.4	**482**	26.9
719	19.7	**689**	22.2	**706**	26.6
704	18.8	**733**	21.4	**632**	26.2
616	18.3	**530**	21.1	**134**	24.3
296	18.2	**458**	20.1	**310**	24.2
442	18.0	**282**	19.7	**715**	22.7
527	17.7	**163**	19.3	**603**	22.6
513	17.2	**123**	19.0	**616**	22.2
602	17.2	**747**	19.0	**674**	22.1
689	16.8	**691**	18.9	**703**	20.8
571	16.7	**502**	18.0	**222**	20.6

Table S4.8. Most influential *m/z* ions in artichoke pectic oligosaccharides (POS) classification according to both substrate and enzyme selected by application of boosted logistic regression (BLR). Importance coefficients were determined by calculating the area under the ROC (Receiver Operating Characteristic) curve.

Known pectic sugars		**Glucanex-Apple**		**Viscozyme-Apple**		**Glucanex-Citrus**	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
121	100.0	**121**	100.0	**411**	100.0	**107**	100.0
282	100.0	**282**	100.0	**426**	100.0	**223**	100.0
411	100.0	**530**	100.0	**428**	100.0	**279**	100.0
426	100.0	**542**	100.0	**442**	100.0	**310**	100.0
428	100.0	**556**	100.0	**457**	100.0	**458**	100.0
442	100.0	**617**	100.0	**513**	100.0	**482**	100.0
457	100.0	**631**	100.0	**516**	100.0	**514**	100.0
513	100.0	**632**	100.0	**527**	100.0	**530**	100.0
516	100.0	**660**	100.0	**530**	100.0	**542**	100.0
527	100.0	**689**	100.0	**531**	100.0	**556**	100.0
530	100.0	**691**	100.0	**542**	100.0	**617**	100.0
531	100.0	**703**	100.0	**544**	100.0	**631**	100.0
544	100.0	**715**	100.0	**556**	100.0	**632**	100.0
603	100.0	**719**	100.0	**603**	100.0	**646**	100.0
616	100.0	**733**	100.0	**616**	100.0	**660**	100.0
617	100.0	**747**	100.0	**617**	100.0	**689**	100.0
630	100.0	**123**	97.8	**630**	100.0	**703**	100.0
646	100.0	**427**	97.7	**631**	100.0	**715**	100.0
647	100.0	**151**	95.6	**632**	100.0	**719**	100.0
658	100.0	**209**	95.6	**646**	100.0	**733**	100.0
660	100.0	**253**	95.6	**647**	100.0	**734**	100.0
689	100.0	**482**	95.6	**658**	100.0	**747**	100.0
691	100.0	**734**	95.3	**660**	100.0	**209**	99.4
704	100.0	**337**	93.5	**689**	100.0	**121**	98.8
706	100.0	**223**	91.3	**691**	100.0	**427**	97.7
715	100.0	**484**	91.3	**703**	100.0	**296**	96.9
719	100.0	**266**	89.1	**704**	100.0	**516**	96.9
720	100.0	**472**	89.1	**706**	100.0	**442**	95.6
733	100.0	**630**	89.1	**715**	100.0	**163**	95.0
734	100.0	**225**	86.9	**719**	100.0	**106**	93.8

Table S4.8. Cont.

Viscozyme-Citrus		Glucanex-PGA		Viscozyme-PGA	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
121	100.0	**107**	100.0	**208**	100.0
252	100.0	**120**	100.0	**427**	100.0
366	100.0	**121**	100.0	**530**	100.0
454	100.0	**223**	100.0	**542**	100.0
502	100.0	**239**	100.0	**556**	100.0
530	100.0	**251**	100.0	**617**	100.0
542	100.0	**252**	100.0	**628**	100.0
543	100.0	**296**	100.0	**631**	100.0
556	100.0	**482**	100.0	**632**	100.0
603	100.0	**530**	100.0	**658**	100.0
613	100.0	**542**	100.0	**660**	100.0
616	100.0	**556**	100.0	**689**	100.0
617	100.0	**617**	100.0	**703**	100.0
630	100.0	**631**	100.0	**715**	100.0
631	100.0	**632**	100.0	**719**	100.0
632	100.0	**660**	100.0	**720**	100.0
646	100.0	**689**	100.0	**731**	100.0
647	100.0	**703**	100.0	**733**	100.0
657	100.0	**715**	100.0	**746**	100.0
660	100.0	**719**	100.0	**747**	100.0
674	100.0	**733**	100.0	**603**	96.5
675	100.0	**747**	100.0	**734**	95.3
689	100.0	**691**	98.4	**151**	89.5
691	100.0	**283**	98.0	**180**	87.8
703	100.0	**427**	97.7	**337**	87.2
704	100.0	**734**	95.3	**106**	86.0
706	100.0	**222**	95.1	**482**	86.0
715	100.0	**225**	95.1	**531**	86.0
719	100.0	**151**	94.3	**327**	84.3
733	100.0	**165**	92.7	**674**	82.6

Table S4.9. Most influential *m/z* ions in artichoke pectic oligosaccharides (POS) classification according to both substrate and enzyme selected by application of multilayer perceptron (MLP). Importance coefficients were determined by the sum of the product of raw input-hidden and hidden-output connection weights.

Known pectic sugars		**Glucanex-Apple**		**Viscozyme-Apple**		**Glucanex-Citrus**	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
530	100.0	**337**	100.0	**266**	100.0	**689**	100.0
603	86.5	**325**	92.1	**546**	69.0	**309**	95.6
181	82.8	**133**	89.6	**516**	64.5	**327**	83.1
647	78.0	**499**	67.7	**472**	56.6	**602**	82.8
617	74.2	**295**	67.2	**252**	52.7	**294**	73.4
355	71.1	**151**	65.1	**123**	52.5	**632**	66.3
720	58.6	**147**	64.7	**137**	49.9	**135**	59.8
558	53.0	**135**	61.9	**281**	48.5	**691**	50.1
514	50.6	**123**	51.3	**715**	48.1	**527**	49.8
311	43.9	**223**	49.6	**151**	45.6	**192**	49.6
660	43.1	**370**	48.8	**251**	45.4	**513**	44.7
731	42.8	**556**	47.4	**674**	43.1	**719**	44.0
207	39.6	**235**	44.7	**181**	42.7	**283**	43.5
542	37.2	**470**	37.0	**311**	40.2	**207**	42.1
209	36.8	**238**	34.0	**238**	39.7	**631**	41.3
427	36.1	**266**	33.5	**443**	39.1	**194**	39.5
528	35.4	**207**	32.0	**502**	37.6	**237**	36.9
178	34.3	**327**	31.8	**310**	36.3	**443**	36.0
249	33.0	**221**	31.7	**121**	36.1	**628**	35.3
646	31.6	**366**	28.4	**222**	35.3	**281**	34.3
691	30.2	**632**	25.2	**177**	35.0	**121**	31.9
631	28.7	**107**	24.3	**208**	33.0	**385**	31.6
194	28.0	**528**	23.0	**527**	31.9	**457**	27.3
733	25.2	**458**	22.6	**531**	31.3	**498**	25.9
163	24.2	**675**	21.3	**530**	30.3	**267**	25.8
310	23.1	**409**	20.2	**413**	30.2	**586**	25.0
613	18.2	**208**	18.0	**631**	28.3	**674**	24.1
147	18.1	**222**	17.9	**280**	27.4	**252**	21.5
715	17.9	**283**	16.2	**703**	26.8	**296**	18.4
443	16.9	**603**	16.1	**107**	26.2	**499**	17.6

Table S4.9. Cont.

Viscozyme-Citrus		Glucanex-PGA		Viscozyme-PGA	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
457	100.0	**225**	100.0	**236**	100.0
309	71.5	**163**	95.0	**747**	97.6
442	69.9	**282**	82.3	**235**	81.9
571	68.7	**134**	75.7	**370**	77.2
616	67.0	**279**	75.3	**133**	75.3
704	65.8	**209**	66.4	**267**	72.0
527	56.2	**249**	58.9	**409**	67.5
674	55.0	**355**	51.9	**221**	64.1
531	48.2	**556**	48.8	**164**	61.6
706	47.7	**542**	45.4	**586**	57.1
180	46.7	**121**	44.6	**252**	56.0
513	44.4	**150**	44.6	**613**	50.3
366	43.0	**699**	37.9	**119**	48.4
470	42.7	**178**	37.7	**657**	47.8
223	42.3	**267**	37.6	**294**	46.8
221	42.3	**192**	36.0	**676**	44.4
192	42.3	**253**	36.0	**179**	43.2
454	40.9	**295**	35.3	**281**	41.7
120	37.8	**731**	35.2	**297**	41.1
295	37.4	**310**	34.4	**715**	41.1
646	34.8	**283**	32.8	**150**	39.1
544	32.9	**237**	32.1	**207**	37.6
165	32.7	**236**	31.9	**472**	35.1
633	32.5	**385**	30.3	**366**	35.0
543	32.4	**511**	30.1	**239**	33.3
337	32.3	**338**	29.6	**237**	33.3
133	31.6	**296**	27.1	**482**	32.5
691	28.0	**164**	27.0	**632**	30.5
720	27.4	**647**	26.5	**733**	29.5
279	26.5	**194**	26.4	**151**	29.4

Table S4.10. Possible chemical structures of other *m/z* ions given by machine learning algorithms, determined by competitive fragment modelling (CFM-ID). These ions correspond to TMSO fragments from pectic di- or trisaccharides.

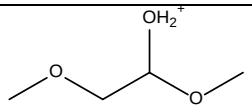
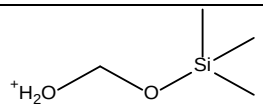
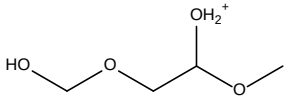
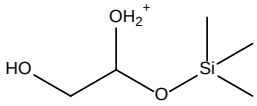
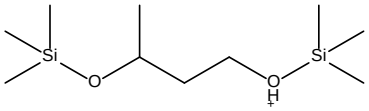
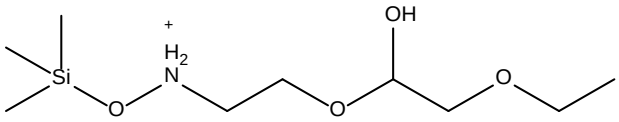
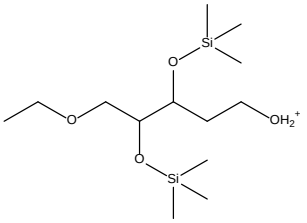
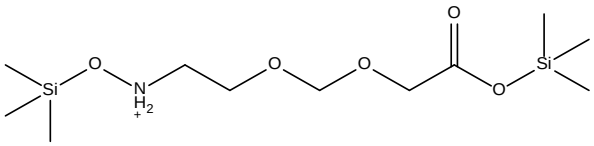
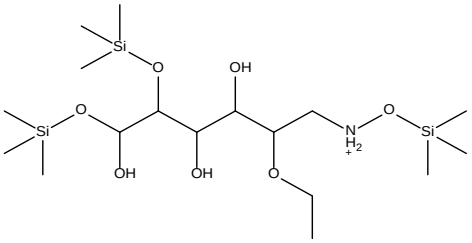
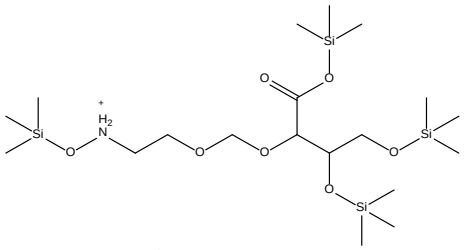
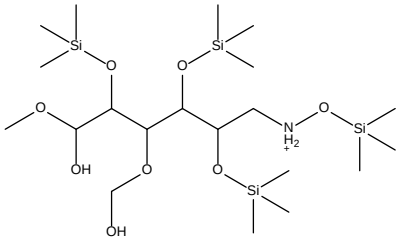
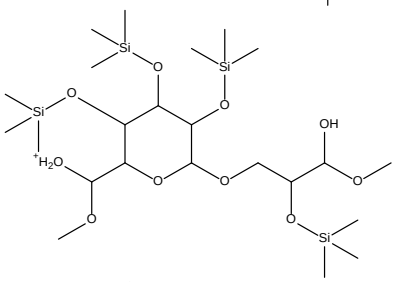
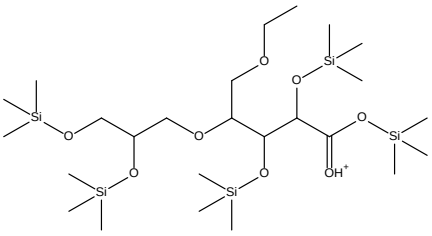
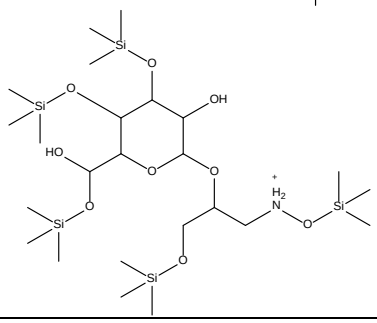
m/z	Derivatized structure	*m/z*	Derivatized structure
107		**121**	
123		**151**	
235		**238**	
309		**310**	
458		**514**	
546		**603**	
629		**646**	

Table S4.10. Cont.

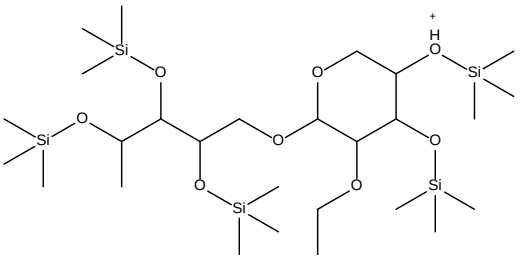
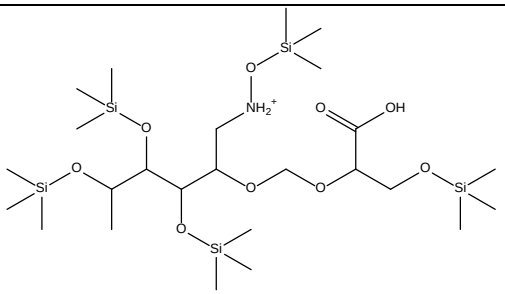
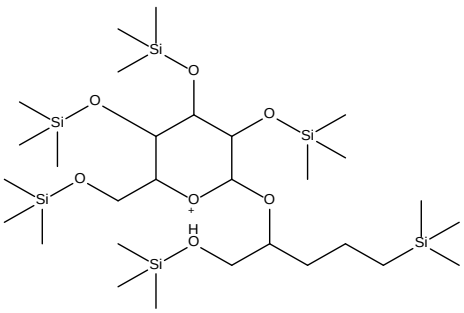
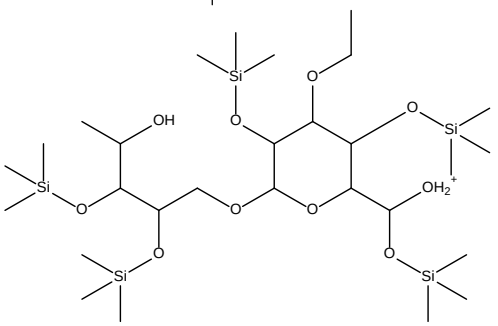
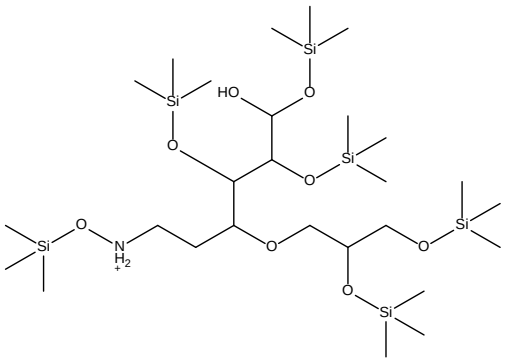
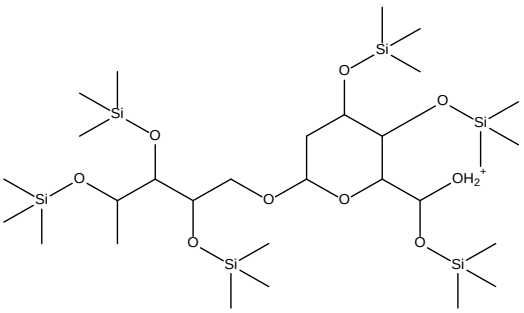
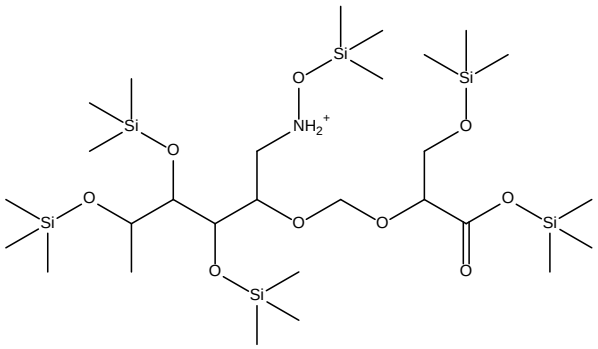
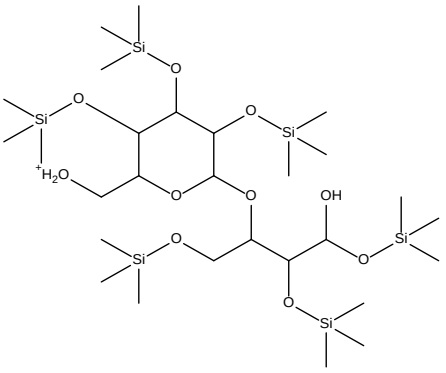
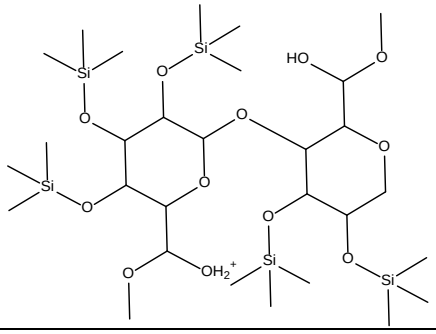
m/z	Derivatized structure	*m/z*	Derivatized structure
657		**660**	
699		**703**	
704		**731**	
732		**733**	
747			

Table S4.11. Sensitivity, specificity, true positives and negatives and balanced accuracy rates (%) for association rule-based classification of POS according to both sample and enzyme source. PGA: polygalacturonic acid.

Parameter	Glucanex-Apple-POS	Viscozyme-Apple-POS	Glucanex-Citrus-POS	Viscozyme-Citrus-POS	Glucanex-PGA-POS	Viscozyme-PGA-POS	Known pectic sugars
Sensitivity	100	100	66.7	100	100	100	95.0
Specificity	98.4	98.3	100	100	100	98.3	100
True positives	83.3	85.7	100	100	100	85.7	100
True negatives	100	100	96.8	100	100	100	97.9
Balanced Accuracy	99.2	99.2	83.3	100	100	99.2	97.5

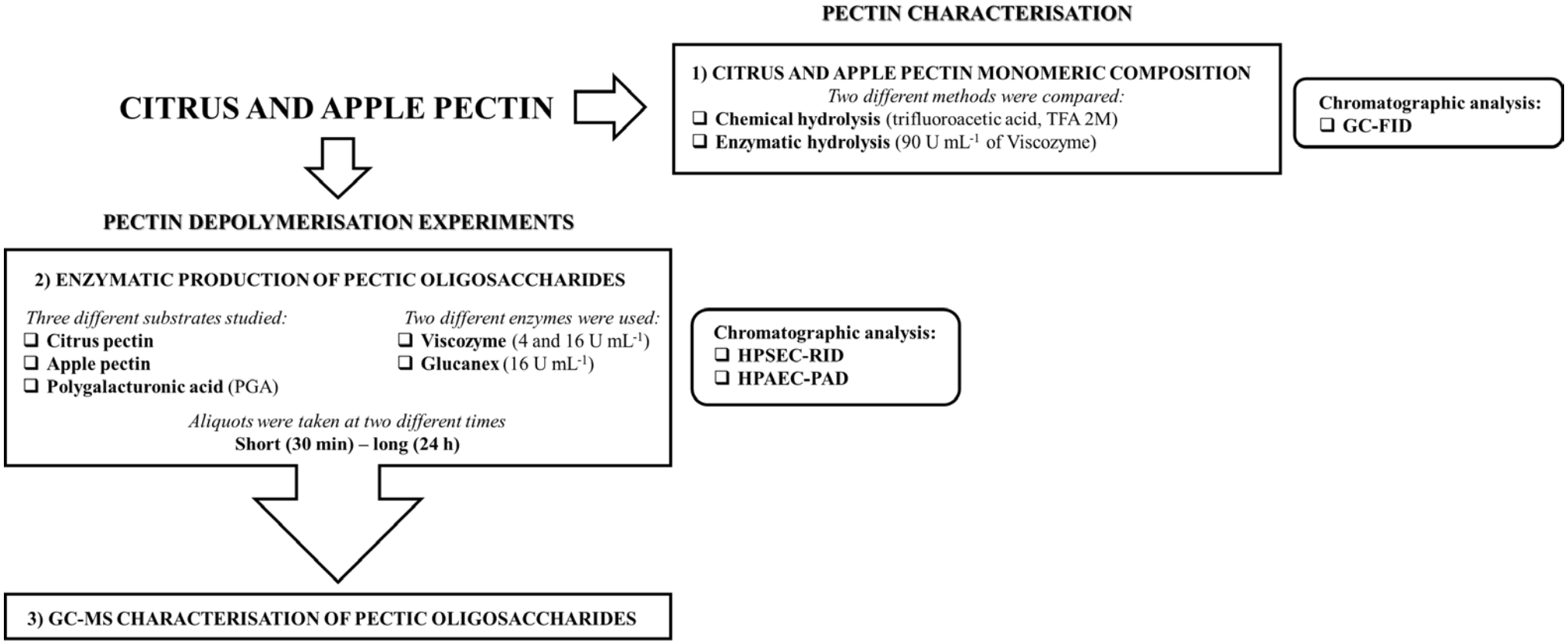


Figure S4.1. Schematic representation of pectin depolymerization experiments and chromatographic analyses.

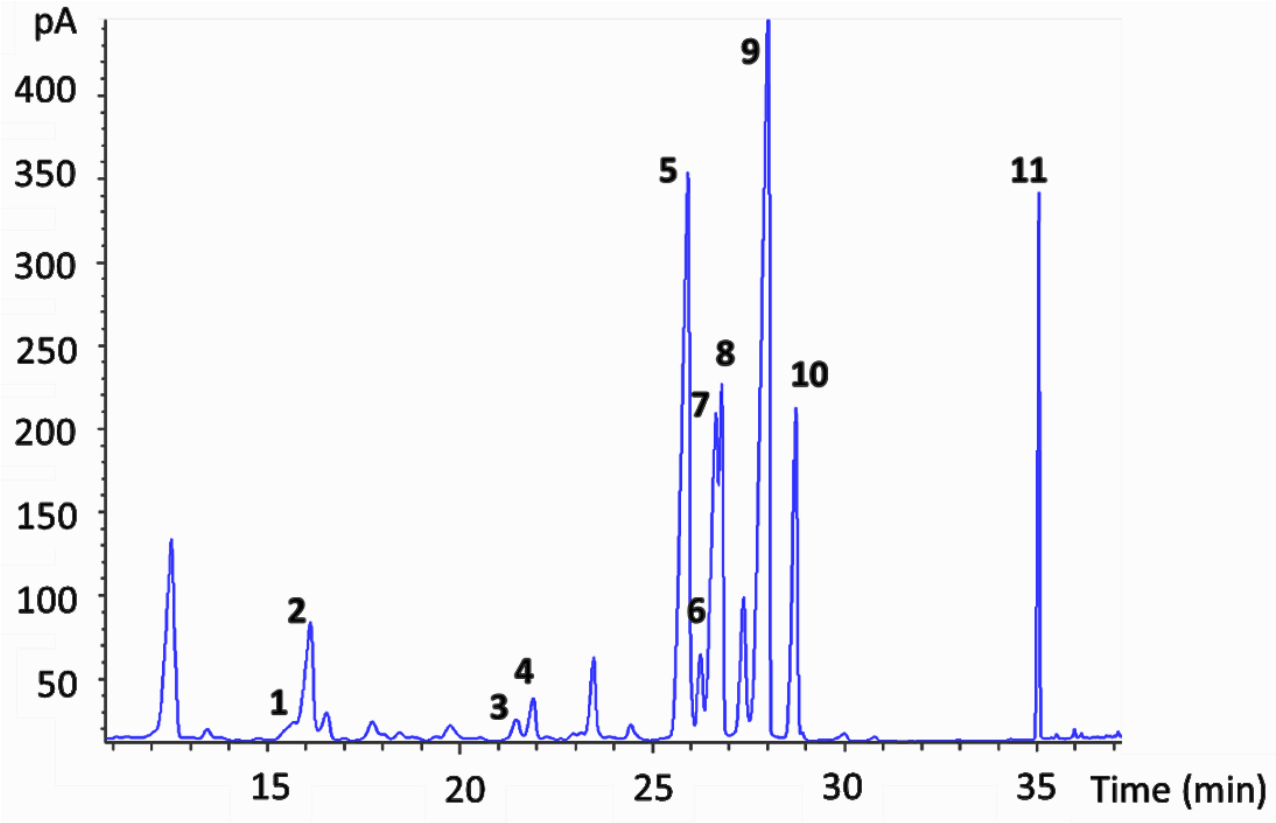


Figure S4.2. Chromatographic profile (GC-FID) of carbohydrates present in hydrolysates from citrus pectin with TFA 2 M during 4 h at 110 ºC. Peaks: (**1**) Xylose, (**2**) Arabinose, (**3, 4**) Rhamnose, (**5**) Galactose, (**6**) Mannose, (**7**) Glucose, (**8**) Galactose+Mannose+Glucose, (**9, 10**) Galacturonic acid, (**11**) Internal standard.

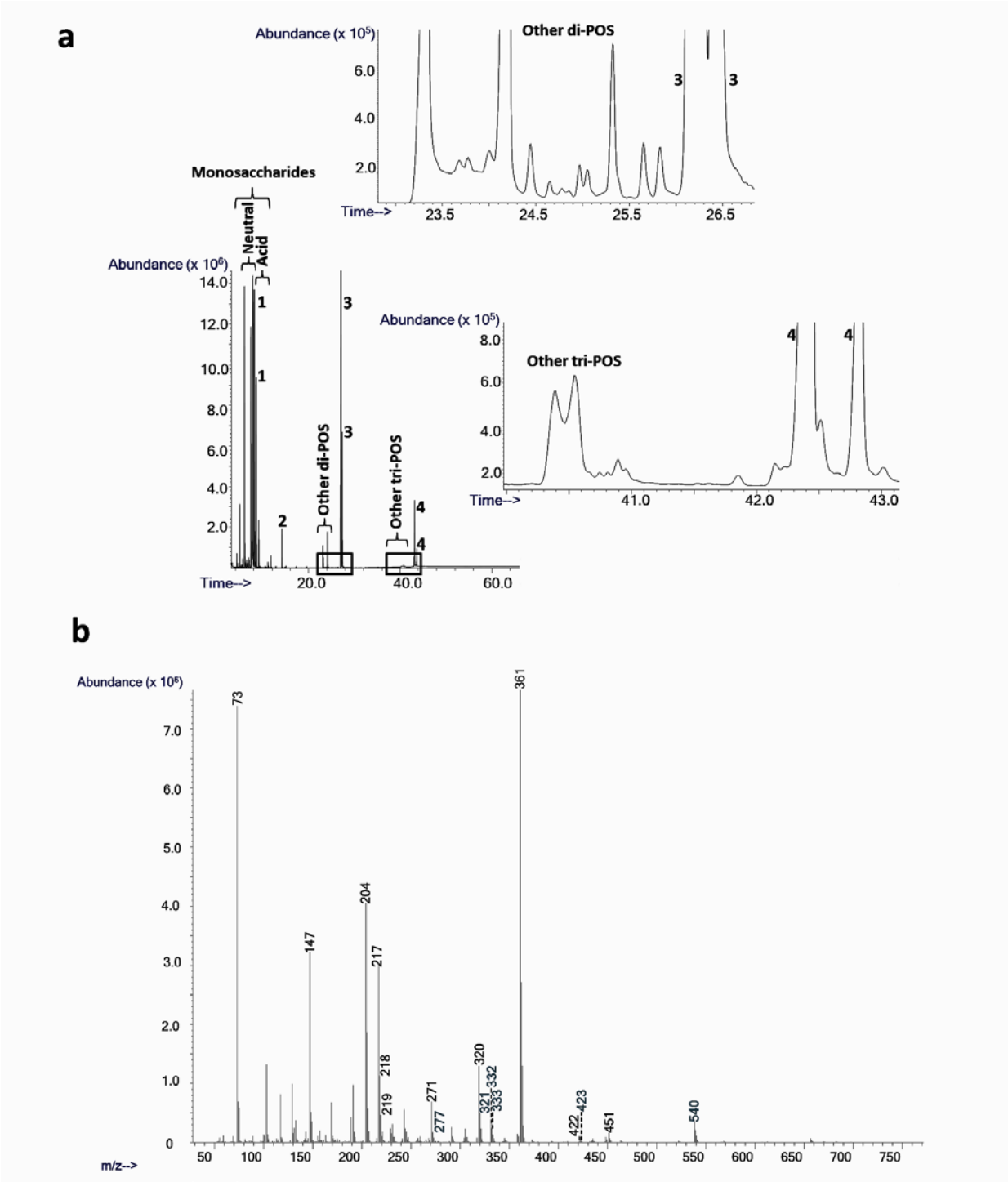


Figure S4.3. a) GC-MS profiles of hydrolysates of citrus pectin obtained by incubation with of Glucanex®200G preparation (16 U mL^{-1}, 30 min at 50 ºC). **a)** Peaks: (1) Galacturonic acid, (2) Internal standard, (3) Digalacturonic acid, (4) Trigalacturonic acid. Di-POS: unknown pectic disaccharides, Tri-POS: unknown pectic trisaccharides. **b)** Characteristic mass spectra of digalacturonic acid (peak 3 above). Galacturonic and uronic acids specific fragments are marked in blue.

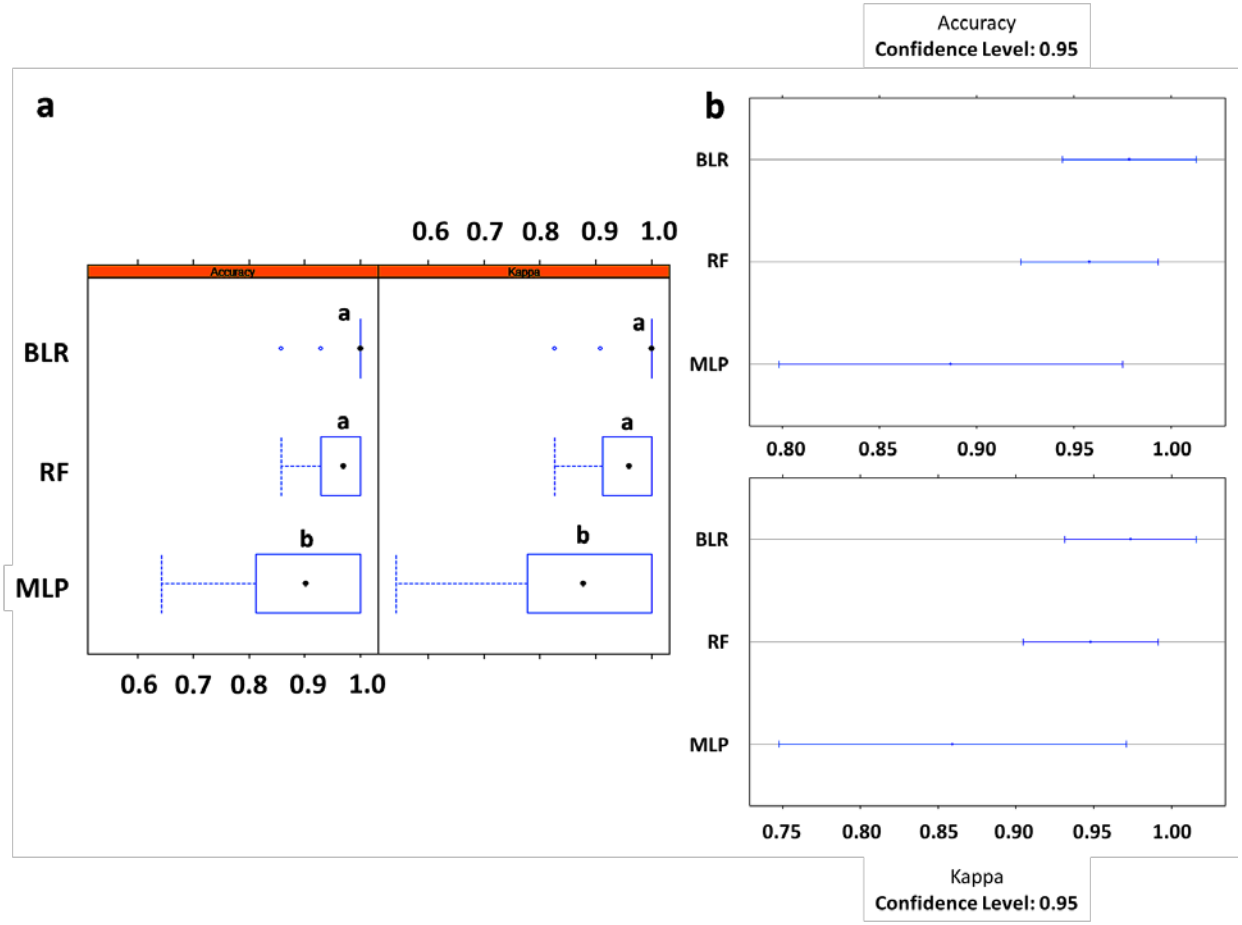


Figure S4.4. Comparative account of the three machine learning algorithms used for POS classification according to both substrate and enzyme selected: boosted logistic regression (BLR), random forest (RF) and multilayer perceptron (MLP). **a)** boxplots of accuracy and kappa values for each models, **b)** accuracy and kappa values obtained at a confidence level of 0.95. Differences between models were calculated via their resampling distributions (number of resamples = 10). [a,b] Statistically significant differences between algorithms.

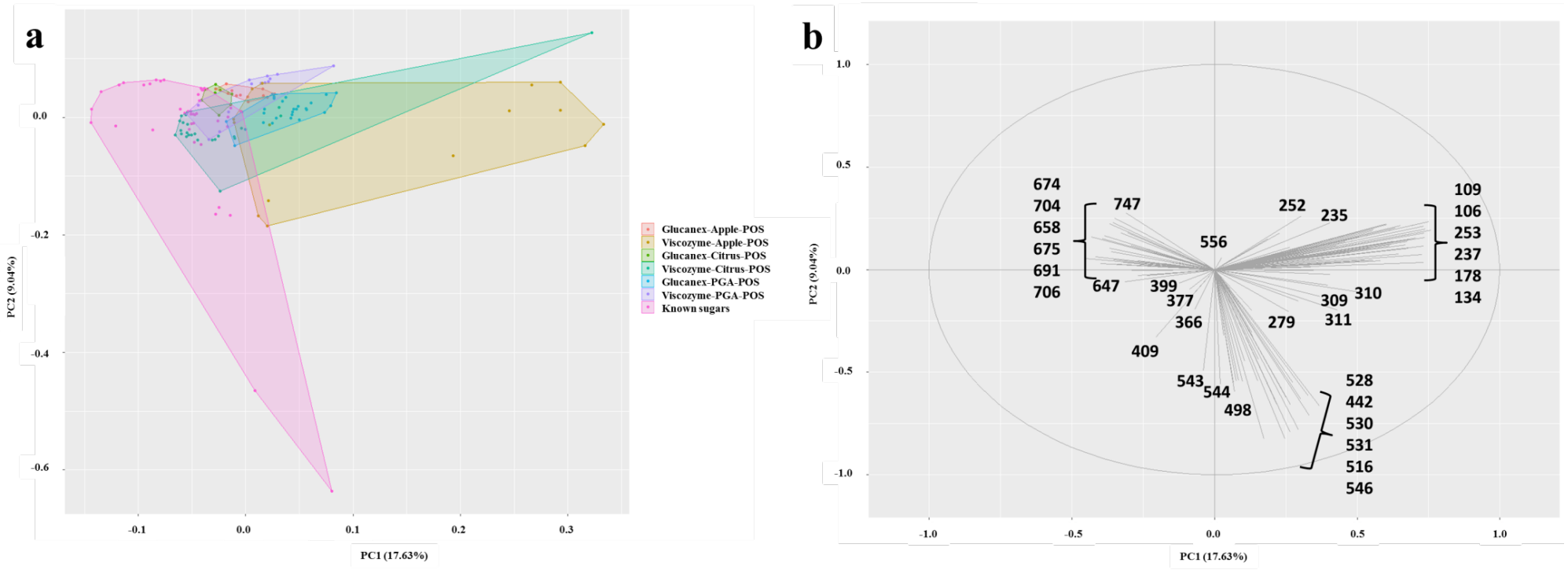


Figure S4.5. PCA scores (**a**) and loadings (**b**) plots of pectic oligosaccharides (POS) obtained with different enzyme preparations (Glucanex and Viscozyme) from studied substrates (polygalacturonic acid, PGA; citrus and apple pectin) considering *m/z* fragments found in their GC-EI-MS spectra.

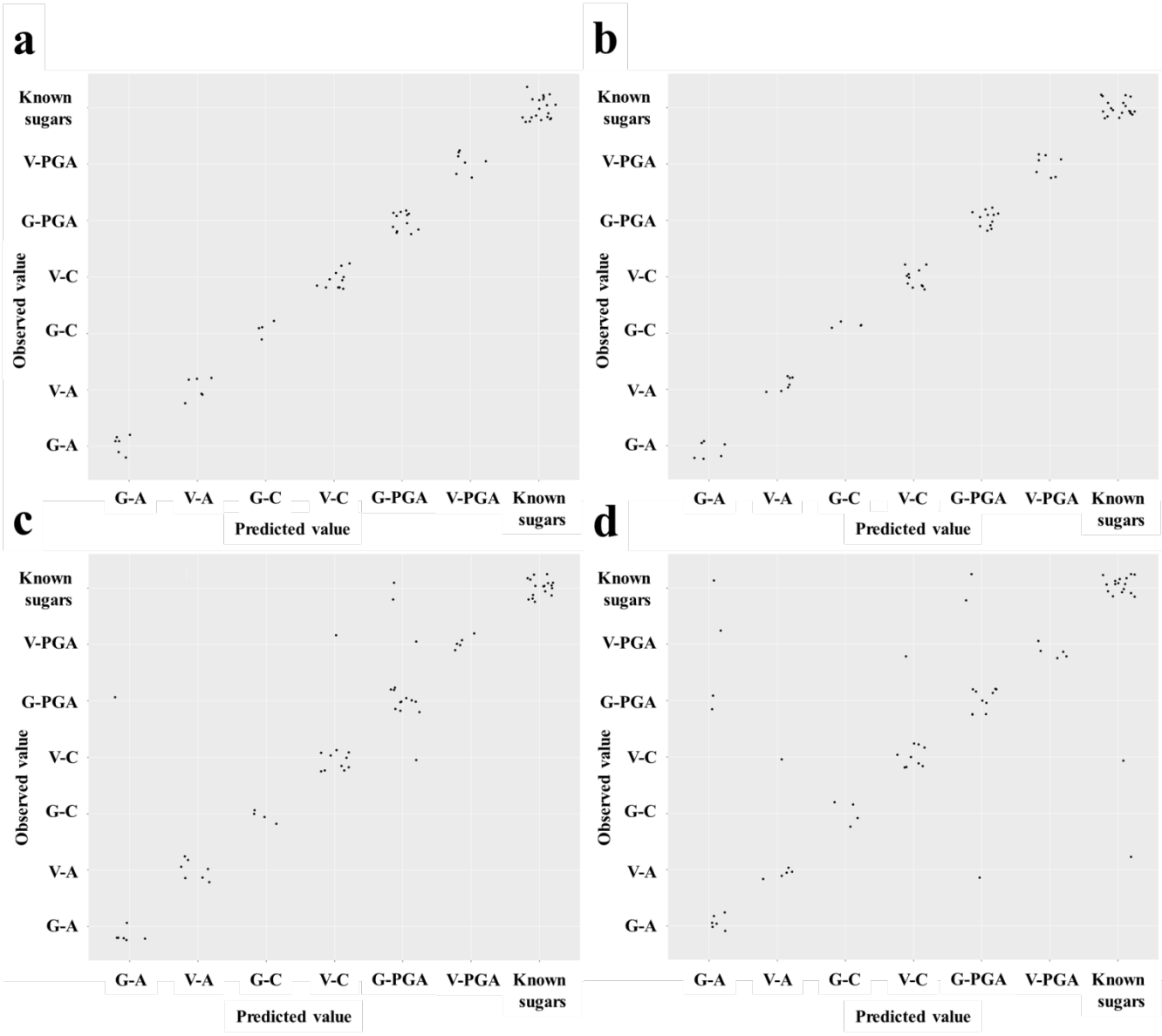


Figure S4.6. Observed vs predicted categories (susbtrate type and enzyme used, test set) of studied pectic oligosaccharides (POS) obtained in different models: **a)** random forest (RF), **b)** bossted logistic regression (BLR), **c)** multilayer perceptros (MLP), **d)** partial least squares-discriminant analysis (PLS-DA). V-: Viscozyme preparation, G-: Glucanex preparation, PGA: polygalacturonic acid, -C: citrus pectin, -A: apple pectin.

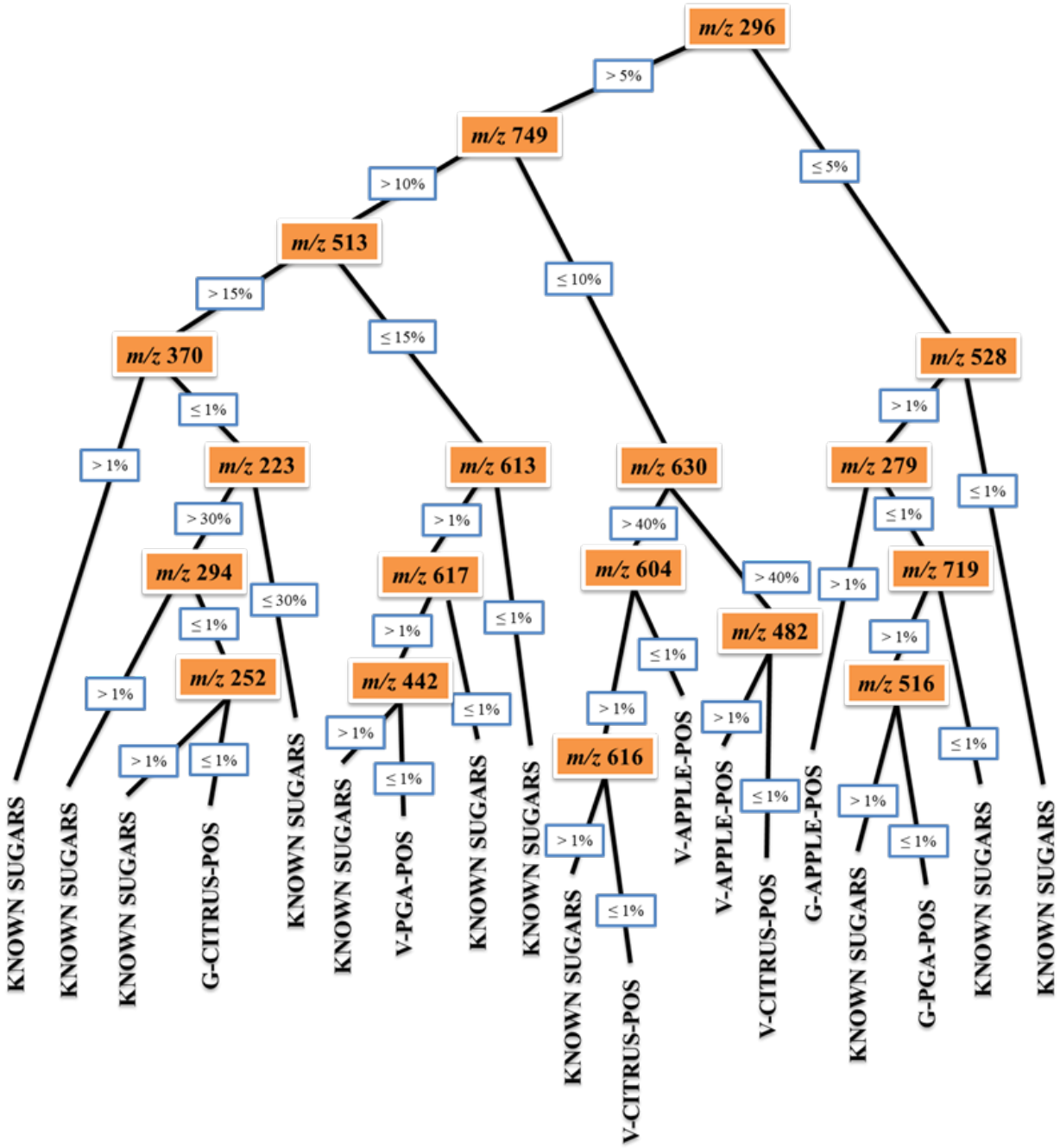


Figure S4.7. Example of a random forest (RF) tree where a set of classification rules are learnt directly from the input data (GC-EI-MS spectra of different pectic oligosaccharides, POS), considering *m/z* ion abundances (%). G-: -: Glucanex preparation, V-: Viscozyme preparation.

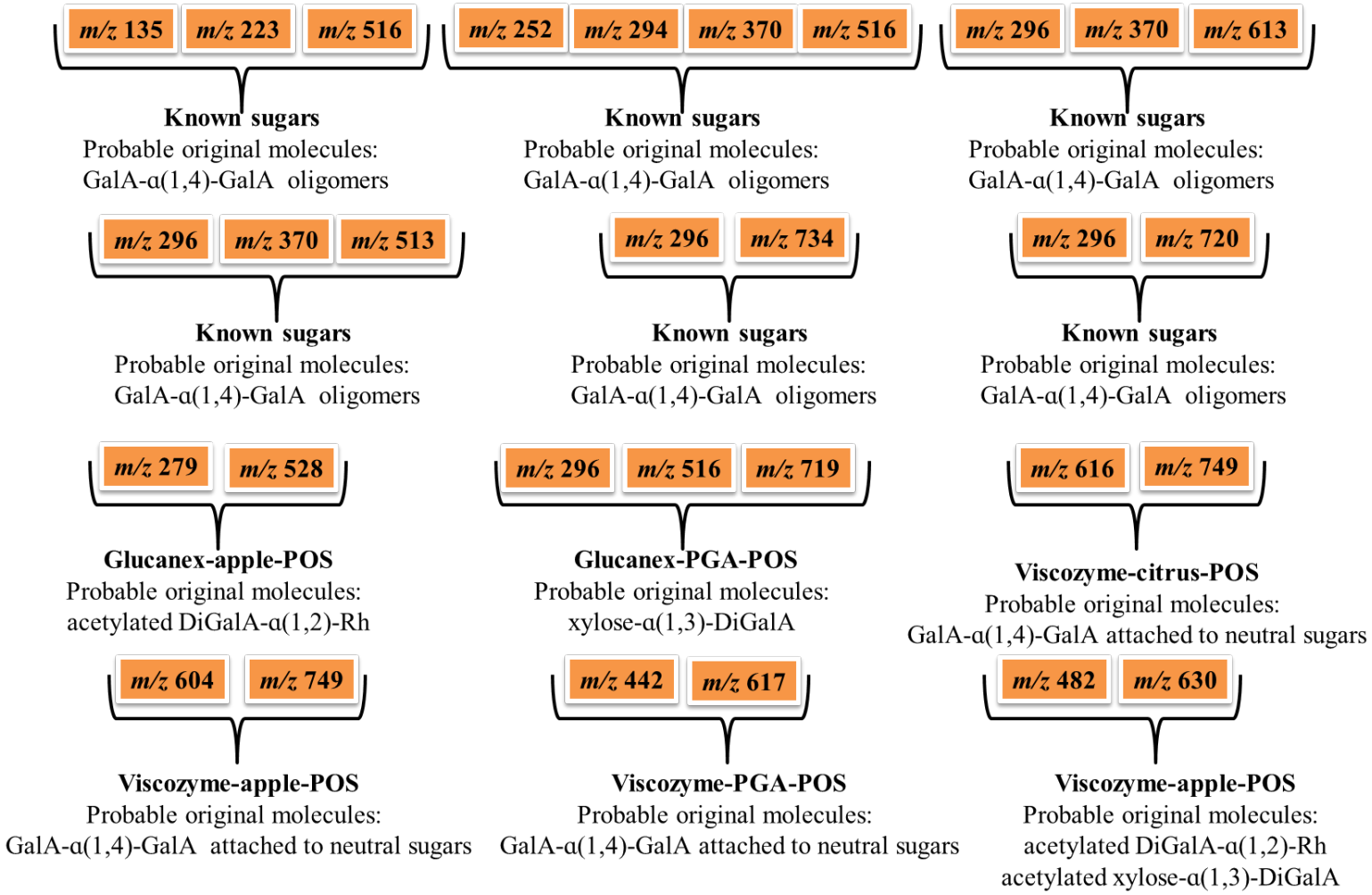


Figure S4.8. Twelve association rules that explained differences in GC-EI-MS spectra of studied pectic oligosaccharides (POS) and possible parent molecules of *m/z* ions. Relationships between ions present in more than one rule are also illustrated. PGA: polygalacturonic acid, GalA: galacturonic acid, Rh: rhamnose.

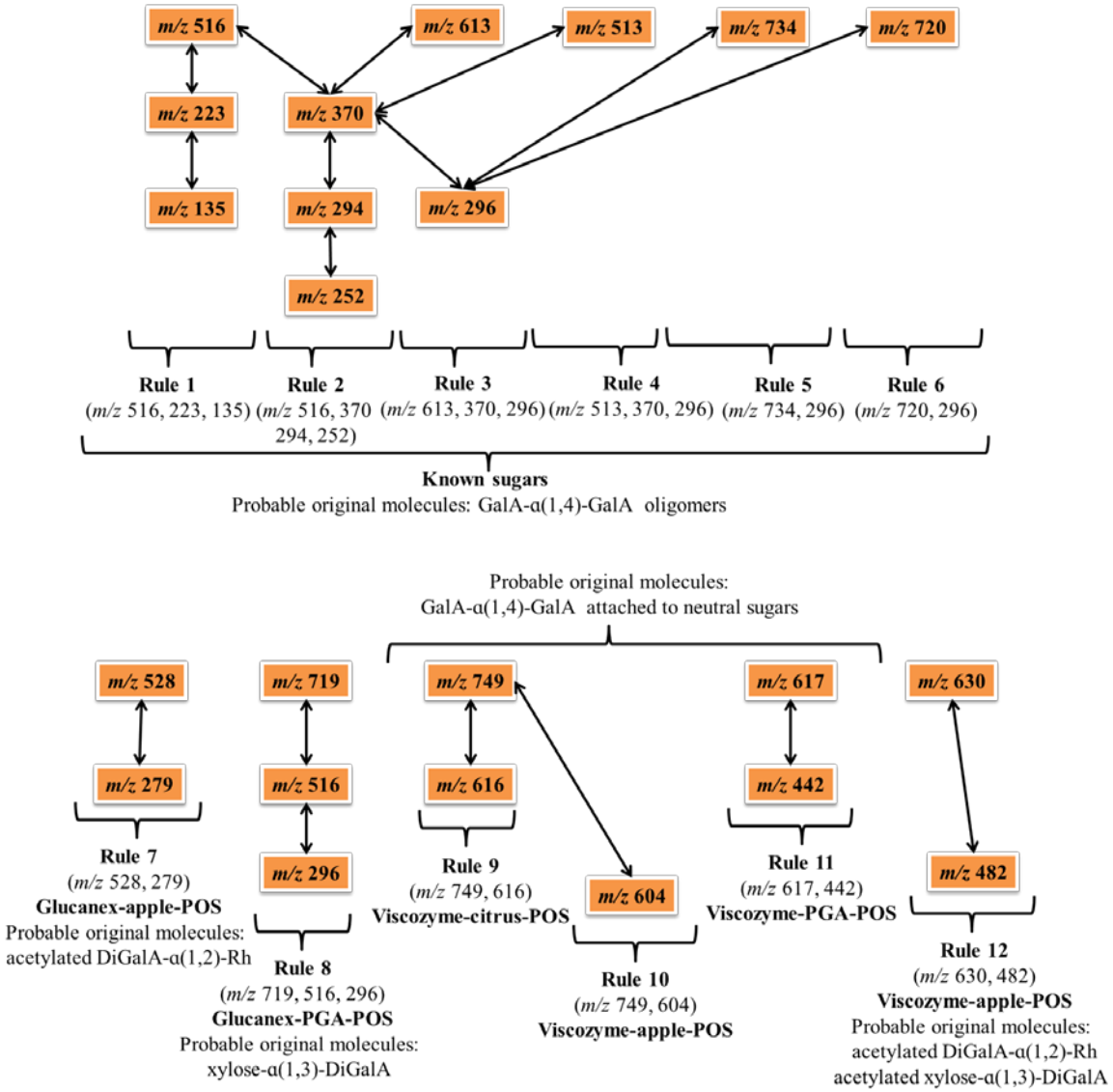


Figure S4.8. Cont.

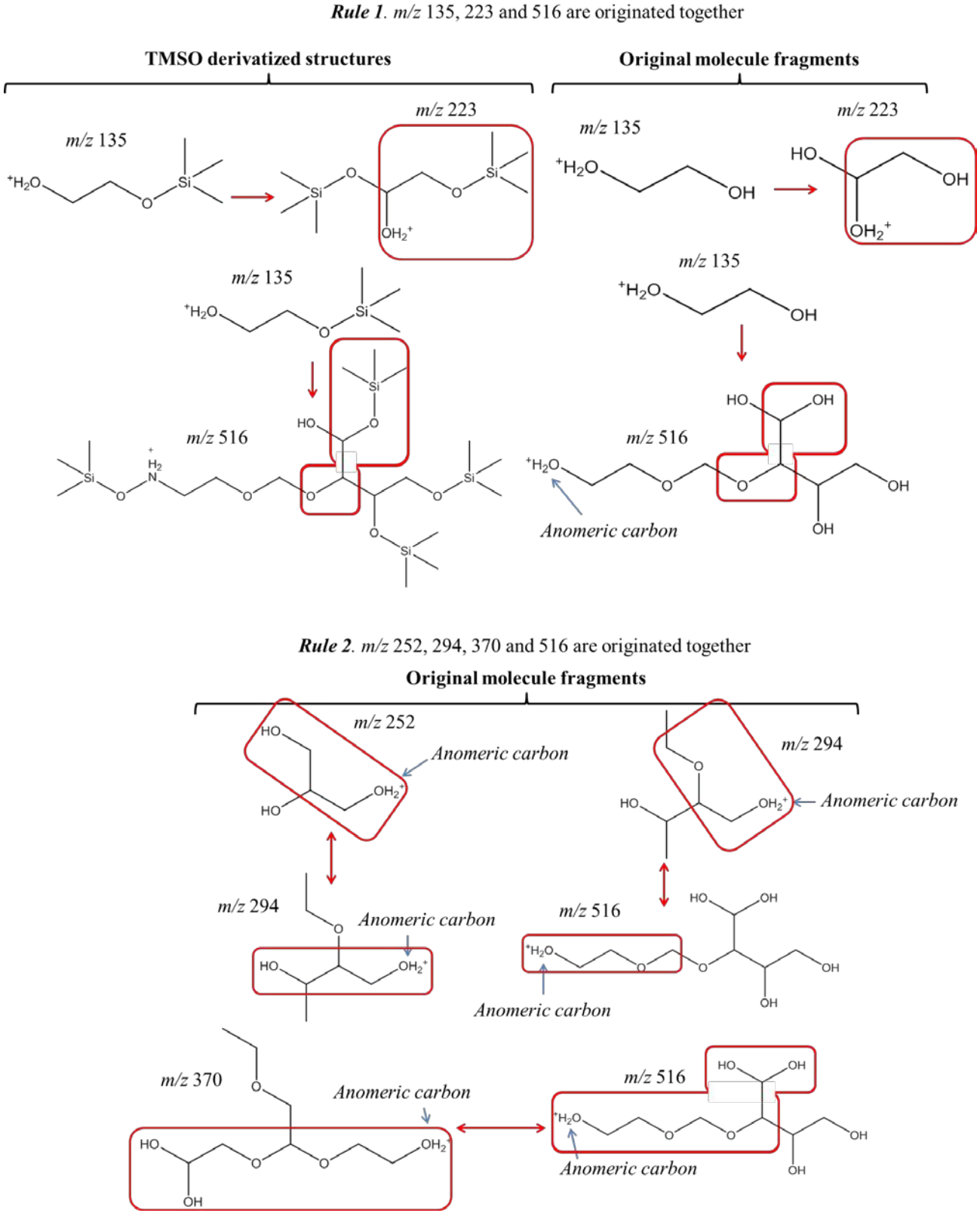
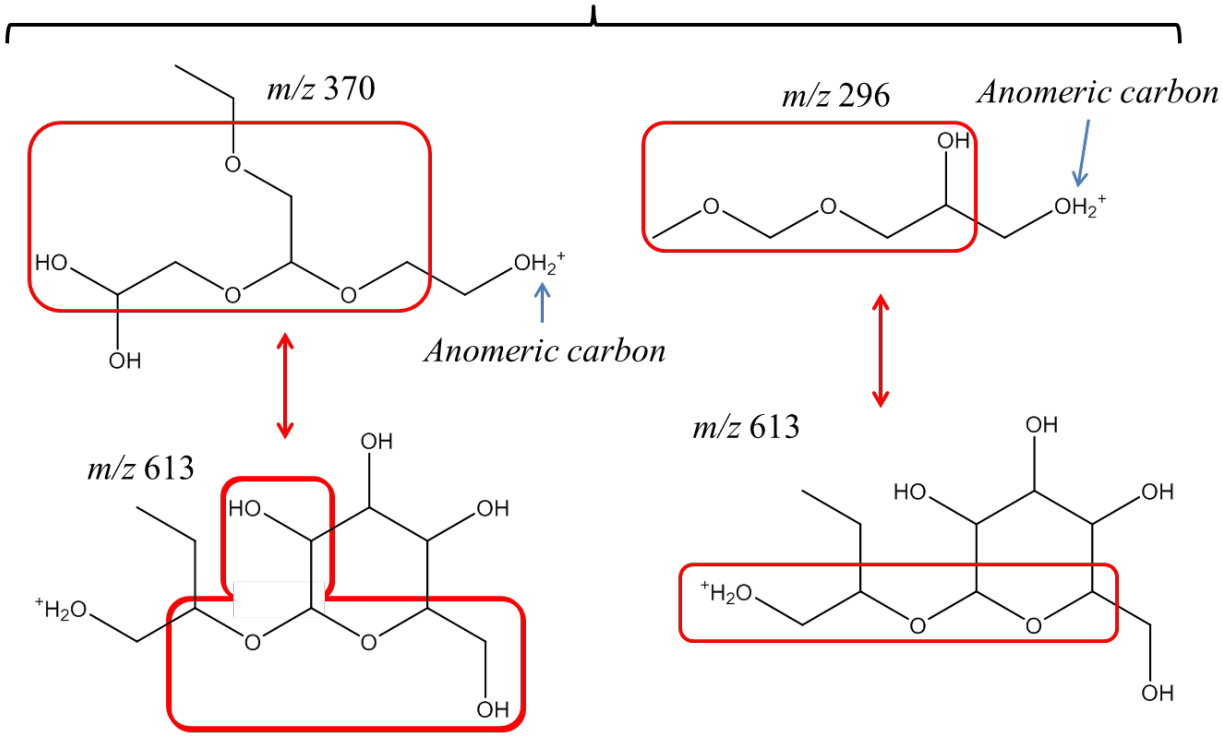


Figure S4.9. Common substructures (marked in red) of those ions formed together in the same association rule. TMSO derivatized structures and original molecule fragments are shown for Rule 1. Common substructures in original molecule fragments are shown for the rest of the Rules.

Rule 3. *m/z* 296, 370 and 613 are originated together



Rule 4. *m/z* 296, 370 and 513 are originated together

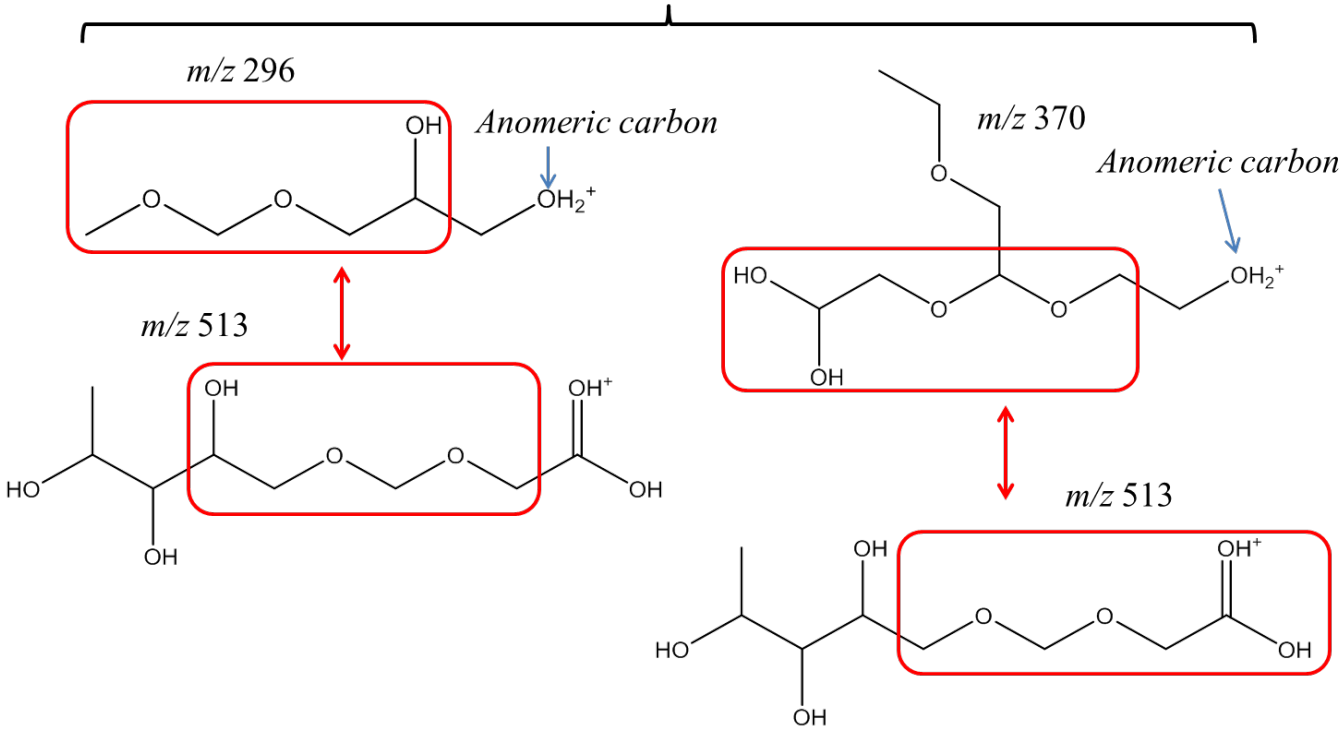
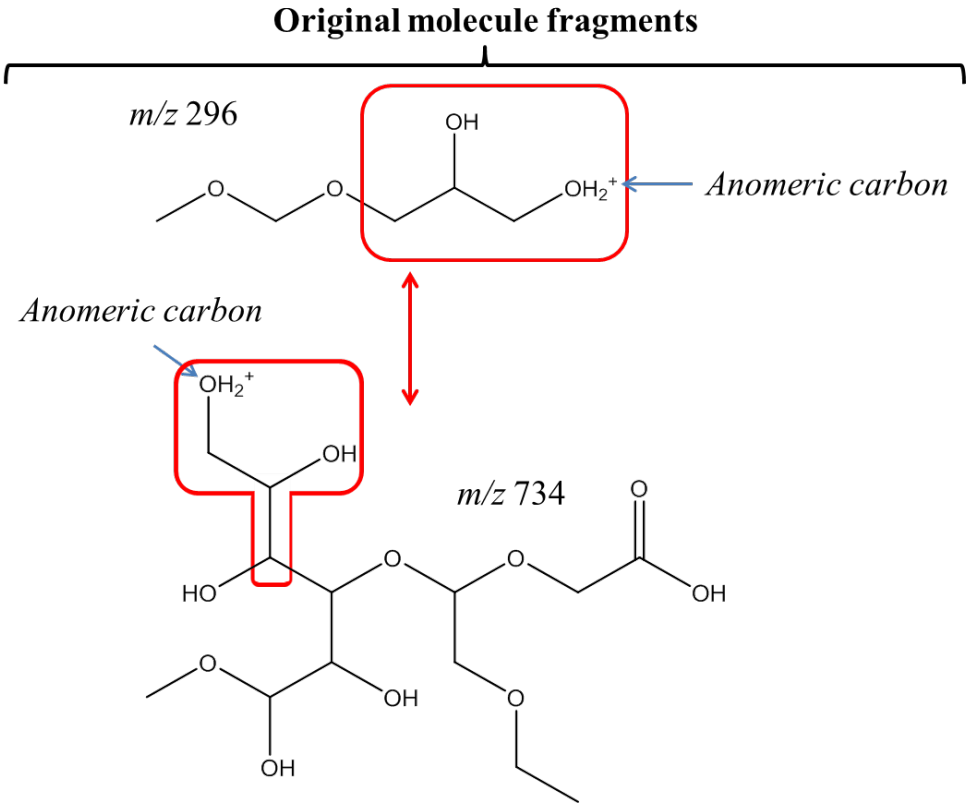


Figure S4.9. Cont.

Rule 5. *m/z* 296 and 734 are originated together



Rule 6. *m/z* 296 and 720 are originated together

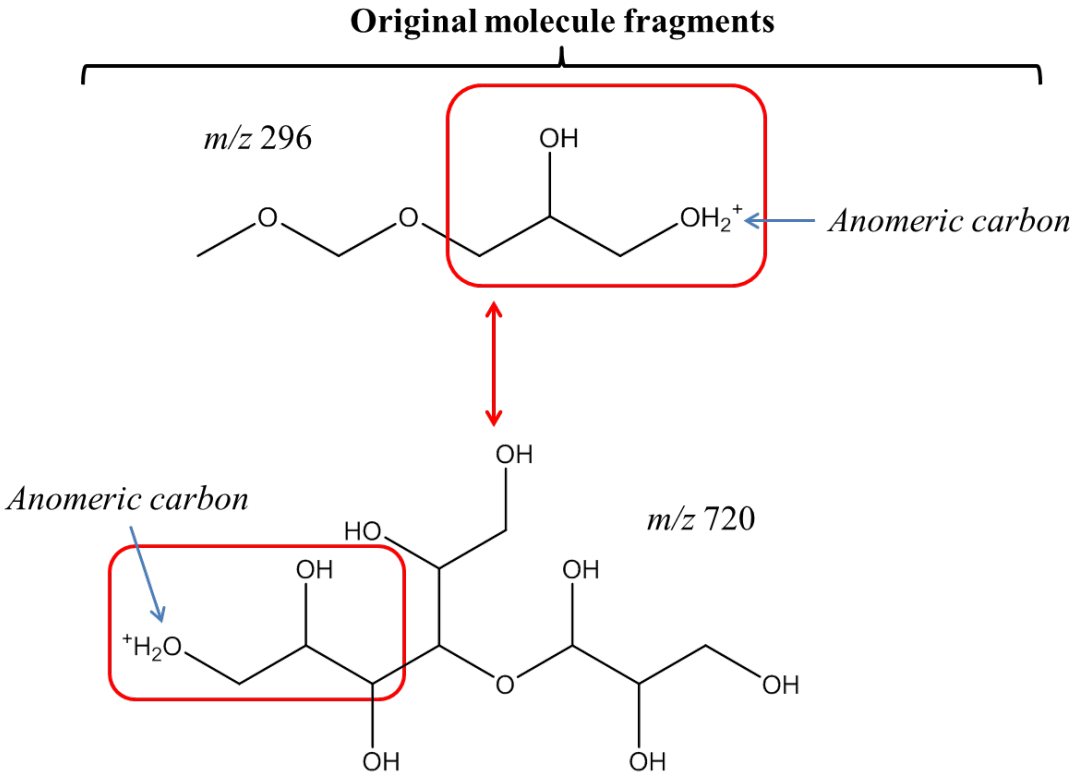
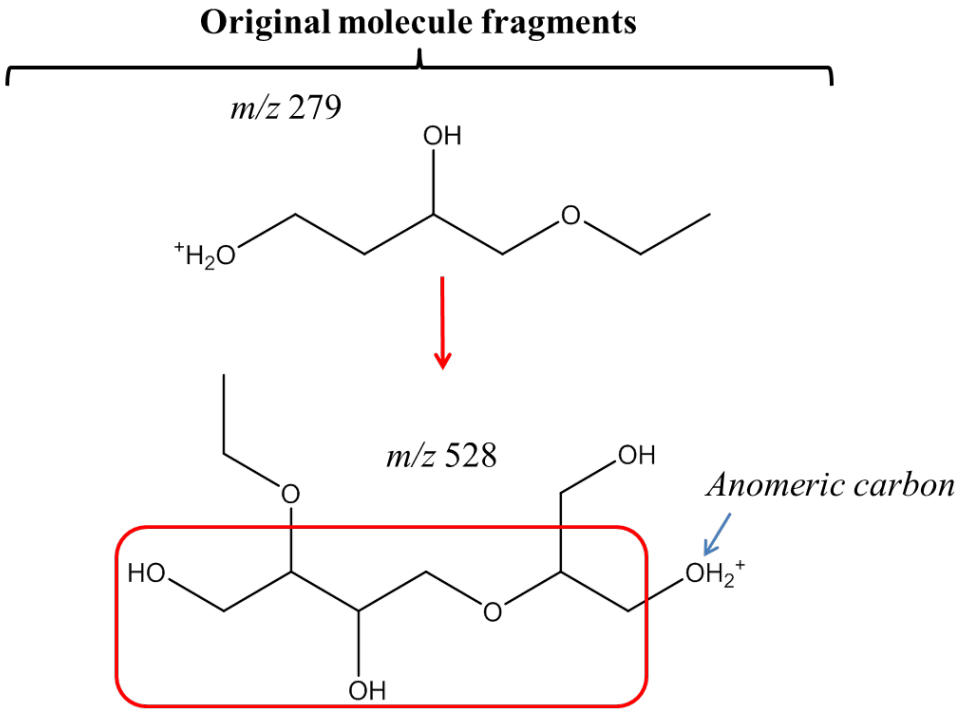


Figure S4.9. Cont.

Rule 7. *m/z* 279 and 528 are originated together



Rule 8. *m/z* 296, 516 and 719 are originated together

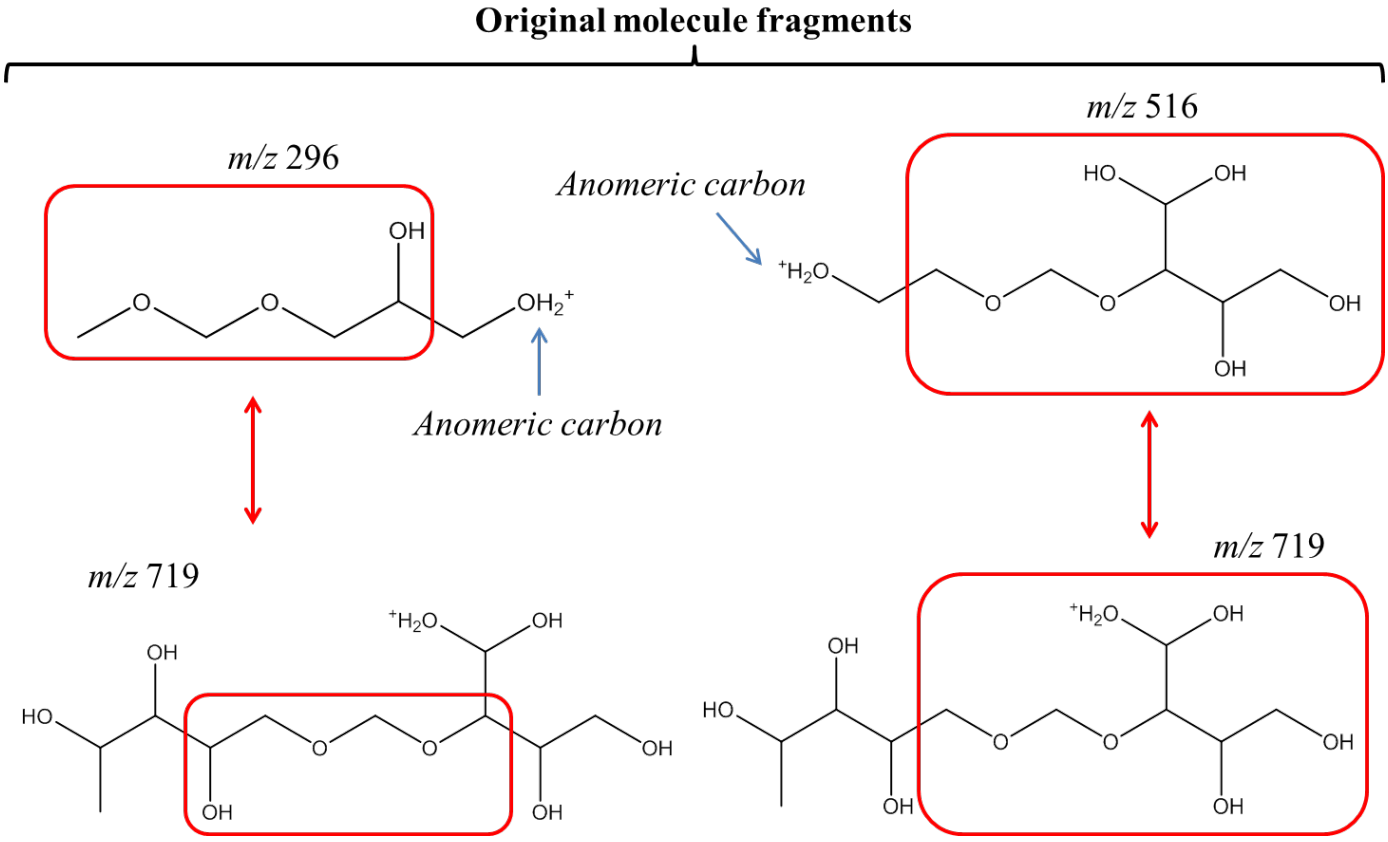
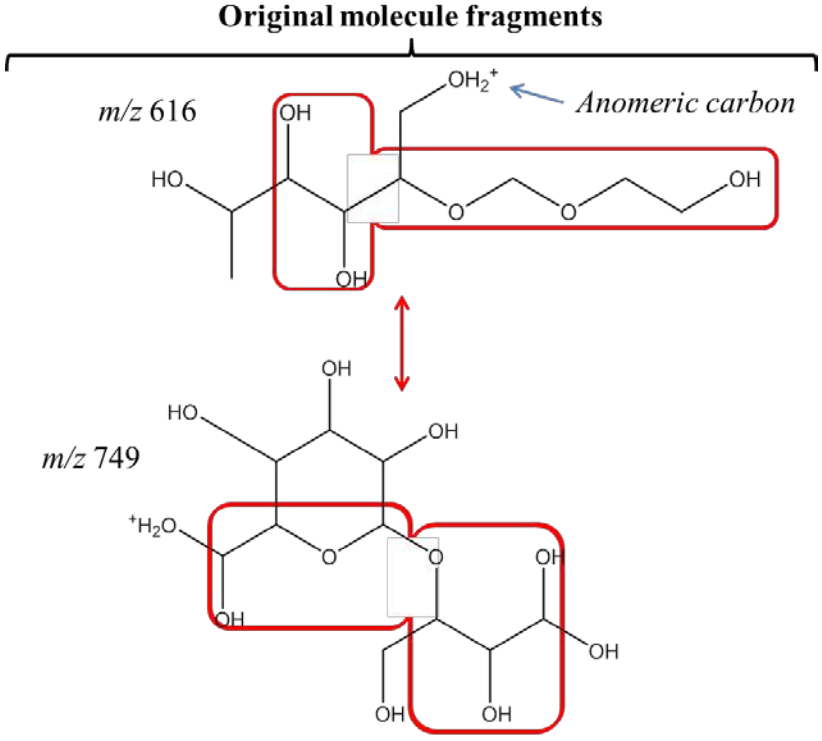


Figure S4.9. Cont.

Rule 9. *m/z* 616 and 749 are originated together



Rule 10. *m/z* 604 and 749 are originated together

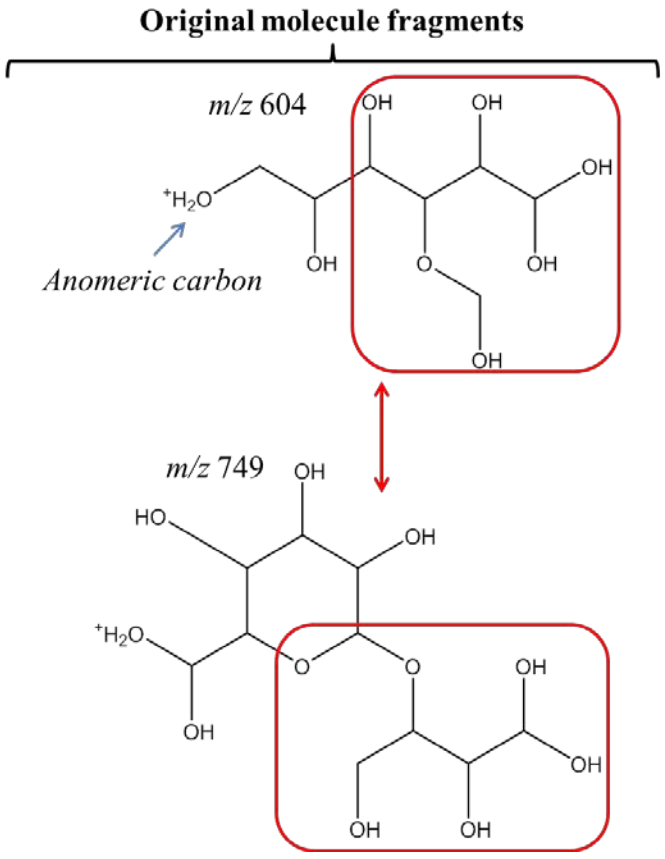
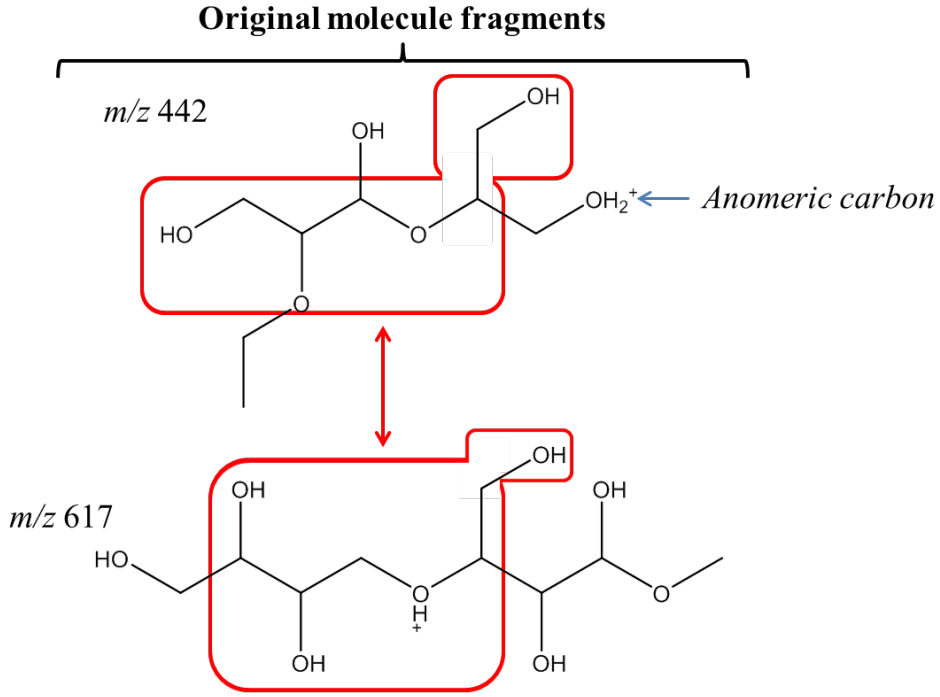


Figure S4.9. Cont.

Rule 11. *m/z* 442 and 617 are originated together



Rule 12. 482 and 630 are originated together

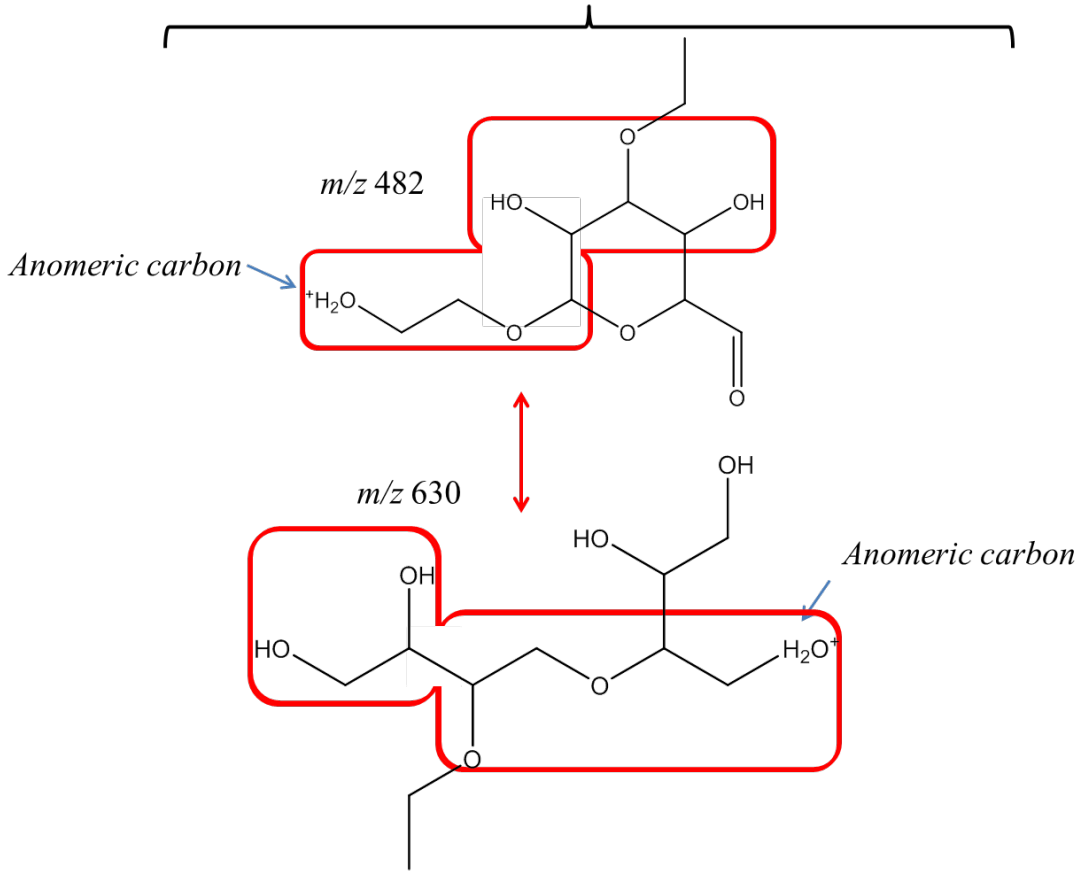


Figure S4.9. Cont.

Annex I.V:

Supplementary material of Article V

11.5. Annex I.V: Supplementary material of Article V

Table S5.1. Characterization of citrus and artichoke pectin and pectic oligosaccharides (**POS**) used as functional ingredients in yogurt elaboration. **DM**: degree of methyl-esterification, $\mathbf{M_w}$: molecular weight. Data mean ± SD.

	Citrus pectin	**Artichoke pectin**	**Citrus POS**	**Artichoke POS**
pH	5.1	6.0	-	-
Dry matter (%)	98.5 ± 0.5	93.3 ± 0.3	-	-
DM (%)	82.0	19.5	-	-
Average mean M_w (kDa)	544 ± 43	222± 5	6.0± 0.4	153.0± 2.6
Distribution of POS [abundance, %]	-	-	-	660 ± 2 [19.7 ± 0.1%]
	-	-	-	85.4 ± 9.9 [19.3 ± 0.1%]
	-	-	9.6 ± 0.7 [50.4 ± 0.3%]	11.0 ± 0.0 [61.0 ± 0.0%]
	-	-	3.8 ± 0.3 [30.2 ± 0.4%]	-
	-	-	0.2 ± 0.0 [10.4 ± 0.1%]	-
Low M_w carbohydrates present in non-hydrolysed mixtures (mg 100 g^{-1} pectin)				
Xylose	3 ± 1	52 ± 3	18 ± 1	92 ± 5.5
Arabinose	24 ± 5	198 ± 18	142 ± 7	409 ± 33
Rhamnose	-	-	-	-
Galactose	45 ± 4	760 ± 19	271 ± 42	1135 ± 212
Galacturonic acid	1635 ± 44	1070 ± 148	23121 ± 53	16244 ± 264
Di-POS	36.2 ± 1.6	6.6 ± 0.6	163.7 ± 20	234.7 ± 9.8
Tri-POS	31.5 ± 1.4	6.7 ± 0.5	366.4 ± 29	2354.0 ± 87.0
Tetra-POS	7.8 ± 0.6	8.8 ± 0.2	-	456.5 ± 6.4

Table S5.2. Number of GC-EI-MS spectra (n = 521) from the carbohydrate fraction of functional yogurts, used for the machine learning classification. ***Myo*-In**: myo-inositol, **Gal**: galactose, **Glc**: glucose, **La**: lactose, **AloLa**: allolactose, **Gabio**: 6-galactobiose, **Ara**: arabinose, **GalA**: galacturonic acid, **GOS**: galacto-oligosaccharides, **POS**: pectic oligosaccharides, **CP**: yogurts with citrus pectin, **AP**: yogurts with artichoke pectin, **CPOS**: yogurts with citrus POS, **APOS**: yogurts with artichoke POS, **-I**: initial oligosaccharides, **-F**: formed oligosaccharides.

				Number of GC-EI-MS spectra							
Yogurt	**Process**	**Time**	**Number of replicates**	***Myo*-In**	**Gal**	**Glc**	**La**	**AloLa**	**GaBio**	**Ara**	**GalA**
Set-style (Control)	**Fermentation (h)**	0	2	2	2*	2*	4	-	-	-	-
		5	2	2	2	2	4	4	4	-	-
	Cold storage (week)	1	2	2	2	2	4	4	4	-	-
Citrus pectin	**Fermentation (h)**	0	2	2	2	2	4	-	-	4	-
		5	2	2	2	2	4	4	4	4	-
	Cold storage (week)	1	2	2	2	2	4	4	4	4	-
Artichoke pectin	**Fermentation (h)**	0	2	2	2	2	4	-	-	4	-
		5	2	2	2	2	4	4	4	4	-
	Cold storage (week)	1	2	2	2	2	4	4	4	4	-
Citrus POS	**Fermentation (h)**	0	2	2	2	2	4	-	-	4	4
		5	2	2	2	2	4	4	4	4	4
	Cold storage (week)	1	2	2	2	2	4	4	4	4	4
Artichoke POS	**Fermentation (h)**	0	2	2	2	2	4	-	-	4	4
		5	2	2	2	2	4	4	4	4	4
	Cold storage (week)	1	2	2	2	2	4	4	4	4	4
Standards	-	-	-	1	2	2	2	-	-	2	2
Total	**-**	**-**	**-**	**31**	**32**	**32**	**62**	**40**	**40**	**50**	**26**

Table S5.2. Cont.

Yogurt	Process	Time	Number of replicates	Number of GC-EI-MS spectra							
				GOS-F	CP-I	CP-F	AP-F**	CPOS-I	CPOS-F	APOS-I	APOS-F
Set-style (Control)	**Fermentation (h)**	0	2	-	-	-	-	-	-	-	-
		5	2	2	-	-	-	-	-	-	-
	Cold storage (week)	1	2	4	-	-	-	-	-	-	-
Citrus pectin	**Fermentation (h)**	0	2	-	2	-	-	-	-	-	-
		5	2	2	2	4	-	-	-	-	-
	Cold storage (week)	1	2	2	2	8	-	-	-	-	-
Artichoke pectin	**Fermentation (h)**	0	2	-	-	-	-	-	-	-	-
		5	2	-	-	-	10	-	-	-	-
	Cold storage (week)	1	2	-	-	-	6	-	-	-	-
Citrus POS	**Fermentation (h)**	0	2	-	-	-	-	16	-	-	-
		5	2	-	-	-	-	16	4	-	-
	Cold storage (week)	1	2	-	-	-	-	16	4	-	-
Artichoke POS	**Fermentation (h)**	0	2	-	-	-	-	-	-	36	-
		5	2	-	-	-	-	-	-	36	12
	Cold storage (week)	1	2	-	-	-	-	-	-	16	8
Standards	-	-	-	-	-	-	-	-	-	-	-
Total	**-**	**-**	**-**	**10**	**6**	**12**	**16**	**48**	**8**	**88**	**20**

* Only the spectra from the first peak was considered because the second peak of galactose and glucose coeluted.

** No initial POS were found in yogurts with artichoke pectin (AP-I).

Table S5.3. Acetic and lactic acid formation during fermentation and cold storage of control and supplemented yogurts.

	Acetic acid (mg 100 g^{-1})						
	0 h	**1 h**	**2 h**	**3 h**	**4 h**	**5 h**	**1 week**
	Fermentation						**Cold storage**
Control	16 ± 3^{e}	18 ± 2^{e}	20 ± 0^{d,e}	21 ± 4^{d,e}	24 ± 2^{d,e}	26 ± 1^{d,e}	35 ± 9^{c,d,e}
Citrus pectin	20 ± 3^{d,e}	23 ± 1^{d,e}	25 ± 3^{d,e}	27 ± 1^{d,e}	29 ± 4^{d,e}	32 ± 5^{d,e}	40 ± 1^{b,c,d}
Citrus POS	25 ± 2^{d,e}	29 ± 1^{d,e}	30 ± 2^{d,e}	31 ± 3^{d,e}	33 ± 0^{d,e}	34 ± 2^{c,d,e}	57 ± 4^{a,b}
Artichoke pectin	23 ± 2^{d,e}	24 ± 0^{d,e}	26 ± 3^{d,e}	28 ± 4^{d,e}	29 ± 1^{d,e}	30 ± 4^{d,e}	55 ± 13^{a,b,c}
Artichoke POS	17 ± 1^{e}	20 ± 3^{d,e}	23 ± 0^{d,e}	27 ± 1^{d,e}	28 ± 3^{d,e}	30 ± 1^{d,e}	67 ± 0^{a}
	Lactic acid (mg 100 g^{-1})						
	0 h	**1 h**	**2 h**	**3 h**	**4 h**	**5 h**	**1 week**
	Fermentation						**Cold storage**
Control	16 ± 3^{h}	26 ± 2^{h}	40 ± 4^{h}	96 ± 12^{h}	465 ± 25^{g}	784 ± 6^{b,c,d,e}	995 ± 106^{a,b}
Citrus pectin	26 ± 1^{h}	30 ± 0^{h}	72 ± 9^{h}	168 ± 12^{h}	525 ± 29^{f,g}	707 ± 27^{c,d,e,f}	951 ± 1^{a,b}
Citrus POS	25 ± 5^{h}	27 ± 2^{h}	30 ± 4^{h}	44 ± 3^{h}	179 ± 1h	441 ± 44^{g}	846 ± 51^{b,c}
Artichoke pectin	26 ± 4^{h}	40 ± 5^{h}	84 ± 9^{h}	180 ± 16^{h}	603 ± 12^{d,e,f,g}	805 ± 10^{b,c,d}	1156 ± 173^{a}
Artichoke POS	27 ± 3^{h}	35 ± 4^{h}	71 ± 8^{h}	158 ± 13^{h}	570 ± 22^{e,f,g}	693 ± 10^{c,d,e,f}	1112 ± 1^{a}

^{a,b,c,d,e,f,g,h} Statistically significant differences between samples.

Table S5.4. Sensitivity, specificity, true positives and negatives and balanced accuracy rates (%) for bagged classification and regression trees (**BCART**) obtained for the test set. ***Myo*-In**: myo-inositol, **Gal**: galactose, **Glc**: glucose, **La**: lactose, **AloLa**: allolactose, **Gabio**: 6-galactobiose, **Ara**: arabinose, **GalA**: galacturonic acid, **GOS**: galacto-oligosaccharides, **POS**: pectic oligosaccharides, **CP**: yogurts with citrus pectin, **AP**: yogurts with artichoke pectin, **CPOS**: yogurts with citrus POS, **APOS**: yogurts with artichoke POS, **-I**: initial oligosaccharides, **-F**: formed oligosaccharides.

Parameter	*Myo*-In	Gal	Glc	La	AloLa	Gabio	Ara	GalA
Sensitivity	100	100	100	86.4	100	100	93.8	100
Specificity	100	99.3	100	100	100	100	100	100
True positives	100	90.0	100	100	100	100	100	100
True negatives	100	100	100	97.8	100	100	99.3	100
Balanced accuracy	100	99.7	100	93.2	100	100	96.0	100

Parameter	GOS-F	CP-I	CP-F	AP-F	CPOS-I	CPOS-F	APOS-I	APOS-F
Sensitivity	100	100	100	100	100	100	96.3	100
Specificity	100	100	100	100	97.9	100	100	99.3
True positives	100	100	100	100	78.6	100	100	83.3
True negatives	100	100	100	100	100	100	99.2	100
Balanced accuracy	100	100	100	100	99.0	100	98.2	99.7

Table S5.5. Sensitivity, specificity, true positives and negatives and balanced accuracy rates (%) for random forest (**RF**) obtained for the test set. ***Myo*-In**: myo-inositol, **Gal**: galactose, **Glc**: glucose, **La**: lactose, **AloLa**: allolactose, **Gabio**: 6-galactobiose, **Ara**: arabinose, **GalA**: galacturonic acid, **GOS**: galacto-oligosaccharides, **POS**: pectic oligosaccharides, **CP**: yogurts with citrus pectin, **AP**: yogurts with artichoke pectin, **CPOS**: yogurts with citrus POS, **APOS**: yogurts with artichoke POS, **-I**: initial oligosaccharides, **-F**: formed oligosaccharides.

Parameter	*Myo*-In	Gal	Glc	La	AloLa	Gabio	Ara	GalA
Sensitivity	100	100	100	100	100	100	93.8	100
Specificity	100	99.3	100	100	100	100	100	100
True positives	100	90.0	100	100	100	100	100	100
True negatives	100	100	100	100	100	100	99.3	100
Balanced accuracy	100	99.7	100	100	100	100	96.9	100

Parameter	GOS-F	CP-I	CP-F	AP-F	CPOS-I	CPOS-F	APOS-I	APOS-F
Sensitivity	75.0	100	100	100	100	100	100	100
Specificity	100	99.4	100	100	100	100	100	100
True positives	100	50.0	100	100	100	100	100	100
True negatives	99.4	100	100	100	100	100	100	100
Balanced accuracy	87.5	99.7	100	100	100	100	100	100

Table S5.6. Most influential *m/z* ions in bagged classification and regression trees (**BCART**) study of yogurt carbohydrates. Importance coefficients were determined by calculating the area under the ROC (Receiver Operating Characteristic) curve. ***Myo*-In**: myo-inositol, **Gal**: galactose, **Glc**: glucose, **La**: lactose, **AloLa**: allolactose, **Gabio**: 6-galactobiose, **Ara**: arabinose, **GalA**: galacturonic acid, **GOS**: galacto-oligosaccharides, **POS**: pectic oligosaccharides, **CP**: yogurts with citrus pectin, **AP**: yogurts with artichoke pectin, **CPOS**: yogurts with citrus POS, **APOS**: yogurts with artichoke POS, **-I**: initial oligosaccharides, **-F**: formed oligosaccharides.

***Myo*-In**		**Gal**		**Glc**		**La**		**Alola**	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
178	100.0	**179**	100.0	**179**	100.0	**179**	100.0	**179**	100.0
179	100.0	**209**	100.0	**209**	100.0	**209**	100.0	**209**	100.0
195	100.0	**238**	100.0	**238**	100.0	**238**	100.0	**238**	100.0
207	100.0	**239**	100.0	**239**	100.0	**251**	100.0	**239**	100.0
209	100.0	**249**	100.0	**267**	100.0	**264**	100.0	**267**	100.0
238	100.0	**251**	100.0	**312**	100.0	**267**	100.0	**297**	100.0
239	100.0	**267**	100.0	**325**	100.0	**295**	100.0	**312**	100.0
251	100.0	**309**	100.0	**327**	100.0	**297**	100.0	**327**	100.0
267	100.0	**311**	100.0	**340**	100.0	**325**	100.0	**353**	100.0
282	100.0	**312**	100.0	**353**	100.0	**327**	100.0	**354**	100.0
297	100.0	**325**	100.0	**354**	100.0	**336**	100.0	**355**	100.0
309	100.0	**327**	100.0	**355**	100.0	**353**	100.0	**368**	100.0
311	100.0	**340**	100.0	**368**	100.0	**354**	100.0	**369**	100.0
325	100.0	**341**	100.0	**369**	100.0	**355**	100.0	**382**	100.0
327	100.0	**353**	100.0	**382**	100.0	**368**	100.0	**383**	100.0
336	100.0	**354**	100.0	**384**	100.0	**369**	100.0	**384**	100.0
339	100.0	**355**	100.0	**385**	100.0	**382**	100.0	**385**	100.0
353	100.0	**367**	100.0	**411**	100.0	**383**	100.0	**411**	100.0
354	100.0	**368**	100.0	**444**	100.0	**384**	100.0	**442**	100.0
355	100.0	**369**	100.0	**458**	100.0	**385**	100.0	**443**	100.0
368	100.0	**382**	100.0	**515**	100.0	**411**	100.0	**444**	100.0
369	100.0	**384**	100.0	**251**	99.7	**425**	100.0	**458**	100.0
382	100.0	**385**	100.0	**311**	99.7	**428**	100.0	**497**	100.0
383	100.0	**411**	100.0	**264**	98.6	**440**	100.0	**513**	100.0
384	100.0	**429**	100.0	**429**	96.6	**444**	100.0	**514**	100.0
385	100.0	**440**	100.0	**195**	96.1	**458**	100.0	**515**	100.0
395	100.0	**442**	100.0	**178**	95.8	**499**	100.0	**530**	100.0
409	100.0	**444**	100.0	**253**	95.8	**511**	100.0	**264**	99.7
411	100.0	**458**	100.0	**341**	95.8	**512**	100.0	**311**	98.7
425	100.0	**499**	100.0	**442**	95.2	**514**	100.0	**499**	97.9

Table S5.6. Cont.

GaBio		Ara		GalA		GOS-F		CP-I	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
179	100.0	**165**	100.0	**179**	100.0	**179**	100.0	**179**	100.0
209	100.0	**179**	100.0	**209**	100.0	**207**	100.0	**209**	100.0
238	100.0	**209**	100.0	**238**	100.0	**395**	100.0	**238**	100.0
239	100.0	**238**	100.0	**239**	100.0	**238**	100.0	**267**	100.0
249	100.0	**239**	100.0	**251**	100.0	**239**	100.0	**327**	100.0
251	100.0	**249**	100.0	**267**	100.0	**267**	100.0	**353**	100.0
267	100.0	**251**	100.0	**327**	100.0	**297**	100.0	**354**	100.0
309	100.0	**267**	100.0	**336**	100.0	**325**	100.0	**355**	100.0
311	100.0	**325**	100.0	**353**	100.0	**327**	100.0	**368**	100.0
312	100.0	**327**	100.0	**354**	100.0	**336**	100.0	**369**	100.0
325	100.0	**340**	100.0	**355**	100.0	**353**	100.0	**382**	100.0
327	100.0	**341**	100.0	**368**	100.0	**354**	100.0	**384**	100.0
340	100.0	**353**	100.0	**369**	100.0	**355**	100.0	**385**	100.0
341	100.0	**354**	100.0	**382**	100.0	**368**	100.0	**411**	100.0
353	100.0	**355**	100.0	**383**	100.0	**369**	100.0	**428**	100.0
354	100.0	**367**	100.0	**384**	100.0	**382**	100.0	**442**	100.0
355	100.0	**368**	100.0	**385**	100.0	**383**	100.0	**444**	100.0
367	100.0	**369**	100.0	**411**	100.0	**384**	100.0	**458**	100.0
368	100.0	**382**	100.0	**425**	100.0	**385**	100.0	**530**	100.0
369	100.0	**383**	100.0	**440**	100.0	**411**	100.0	**426**	99.5
382	100.0	**384**	100.0	**442**	100.0	**425**	100.0	**325**	97.1
384	100.0	**385**	100.0	**444**	100.0	**442**	100.0	**429**	96.6
385	100.0	**411**	100.0	**458**	100.0	**444**	100.0	**195**	96.1
411	100.0	**440**	100.0	**497**	100.0	**458**	100.0	**178**	95.8
428	100.0	**444**	100.0	**514**	100.0	**497**	100.0	**253**	95.8
429	100.0	**458**	100.0	**515**	100.0	**499**	100.0	**341**	95.8
440	100.0	**253**	98.3	**529**	100.0	**514**	100.0	**281**	94.9
444	100.0	**429**	98.3	**530**	100.0	**515**	100.0	**427**	94.5
458	100.0	**195**	96.1	**207**	98.9	**530**	100.0	**297**	93.5
514	100.0	**529**	96.0	**409**	98.9	**195**	99.3	**383**	92.8

Table S5.6. Cont.

CP-F		AP-F		CPOS-I		CPOS-F		APOS-I	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
179	100.0	**179**	100.0	**178**	100.0	**165**	100.0	**179**	100.0
209	100.0	**209**	100.0	**179**	100.0	**179**	100.0	**209**	100.0
238	100.0	**238**	100.0	**209**	100.0	**238**	100.0	**238**	100.0
267	100.0	**251**	100.0	**238**	100.0	**239**	100.0	**264**	100.0
280	100.0	**267**	100.0	**239**	100.0	**249**	100.0	**267**	100.0
327	100.0	**309**	100.0	**267**	100.0	**251**	100.0	**295**	100.0
353	100.0	**311**	100.0	**311**	100.0	**267**	100.0	**309**	100.0
354	100.0	**325**	100.0	**312**	100.0	**297**	100.0	**327**	100.0
355	100.0	**327**	100.0	**327**	100.0	**312**	100.0	**353**	100.0
368	100.0	**336**	100.0	**341**	100.0	**325**	100.0	**354**	100.0
369	100.0	**339**	100.0	**353**	100.0	**327**	100.0	**355**	100.0
382	100.0	**340**	100.0	**354**	100.0	**340**	100.0	**368**	100.0
384	100.0	**353**	100.0	**355**	100.0	**341**	100.0	**369**	100.0
385	100.0	**354**	100.0	**368**	100.0	**353**	100.0	**382**	100.0
411	100.0	**355**	100.0	**369**	100.0	**354**	100.0	**384**	100.0
428	100.0	**368**	100.0	**382**	100.0	**355**	100.0	**385**	100.0
440	100.0	**369**	100.0	**384**	100.0	**367**	100.0	**411**	100.0
442	100.0	**382**	100.0	**385**	100.0	**368**	100.0	**428**	100.0
443	100.0	**383**	100.0	**411**	100.0	**369**	100.0	**444**	100.0
444	100.0	**384**	100.0	**443**	100.0	**382**	100.0	**458**	100.0
458	100.0	**385**	100.0	**444**	100.0	**383**	100.0	**511**	100.0
325	97.1	**409**	100.0	**458**	100.0	**384**	100.0	**512**	100.0
426	96.8	**411**	100.0	**163**	99.4	**385**	100.0	**514**	100.0
429	96.6	**425**	100.0	**325**	97.1	**411**	100.0	**336**	99.3
195	96.1	**428**	100.0	**429**	96.6	**440**	100.0	**325**	97.1
337	96.1	**442**	100.0	**195**	96.1	**442**	100.0	**429**	96.6
178	95.8	**444**	100.0	**253**	95.8	**443**	100.0	**383**	96.5
253	95.8	**458**	100.0	**281**	94.9	**444**	100.0	**195**	96.1
341	95.8	**499**	100.0	**297**	93.5	**458**	100.0	**178**	95.8
281	94.9	**511**	100.0	**383**	92.8	**497**	100.0	**253**	95.8

Table S5.6. Cont.

	APOS-F
m/z	**Importance**
179	100.0
209	100.0
238	100.0
251	100.0
264	100.0
267	100.0
281	100.0
239	100.0
309	100.0
325	100.0
327	100.0
341	100.0
353	100.0
354	100.0
355	100.0
368	100.0
369	100.0
382	100.0
383	100.0
384	100.0
385	100.0
411	100.0
428	100.0
440	100.0
442	100.0
443	100.0
444	100.0
456	100.0
458	100.0
511	100.0

Table S5.7. Most influential *m/z* ions in random forest (**RF**) study of yogurt carbohydrates. Importance coefficients were determined by the permutation of the out of-bag-error. ***Myo*-In**: myo-inositol, **Gal**: galactose, **Glc**: glucose, **La**: lactose, **AloLa**: allolactose, **Gabio**: 6-galactobiose, **Ara**: arabinose, **GalA**: galacturonic acid, **GOS**: galacto-oligosaccharides, **POS**: pectic oligosaccharides**, CP**: yogurts with citrus pectin, **AP**: yogurts with artichoke pectin, **CPOS**: yogurts with citrus POS, **APOS**: yogurts with artichoke POS, **-I**: initial oligosaccharides, **-F**: formed oligosaccharides.

***Myo*-In**		**Gal**		**Glc**		**La**		**Alola**	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
530	63.8	**355**	59.1	**530**	64.3	**384**	93.8	**382**	69.5
355	57.4	**384**	54.3	**355**	56.3	**530**	64.8	**322**	68.5
382	54.9	**382**	53.5	**384**	54.9	**382**	52.2	**336**	63.6
442	54.8	**327**	47.1	**382**	52.0	**383**	43.0	**442**	48.6
384	53.3	**251**	44.8	**442**	52.0	**515**	38.6	**384**	46.2
385	46.9	**383**	42.4	**383**	45.9	**442**	38.4	**514**	37.0
428	44.2	**253**	42.3	**327**	43.0	**339**	35.2	**339**	35.6
515	43.8	**385**	41.0	**411**	42.5	**355**	35.1	**327**	35.0
383	43.7	**249**	38.7	**239**	40.9	**426**	33.6	**383**	34.5
514	43.3	**411**	38.5	**514**	40.8	**312**	31.8	**354**	32.6
411	41.3	**264**	38.2	**499**	40.6	**514**	30.3	**411**	32.4
512	41.1	**239**	35.9	**354**	39.1	**411**	28.8	**426**	32.2
264	40.8	**310**	35.4	**165**	37.5	**444**	28.2	**515**	31.7
528	39.9	**339**	35.2	**311**	36.8	**370**	28.2	**312**	31.4
327	37.2	**354**	34.0	**312**	36.4	**354**	27.8	**165**	30.1
311	37.2	**311**	33.8	**253**	36.2	**340**	27.4	**512**	30.1
165	35.6	**442**	31.2	**385**	35.8	**327**	27.2	**238**	29.7
339	34.8	**515**	31.0	**264**	35.3	**368**	26.5	**444**	29.4
529	33.3	**165**	30.9	**515**	34.7	**369**	26.3	**385**	29.1
426	32.0	**367**	30.8	**310**	34.7	**294**	25.7	**325**	28.4
310	31.3	**444**	30.4	**249**	34.7	**311**	25.6	**264**	27.7
239	30.8	**294**	30.1	**339**	33.6	**385**	25.4	**428**	27.3
312	30.8	**179**	29.9	**444**	33.5	**249**	25.0	**353**	27.2
427	30.4	**542**	29.6	**251**	33.0	**238**	24.7	**179**	26.4
179	30.0	**499**	29.4	**512**	33.0	**325**	24.7	**239**	26.3
251	29.6	**312**	29.4	**179**	31.8	**165**	24.6	**311**	25.1
249	29.6	**296**	28.9	**529**	29.5	**322**	23.6	**443**	24.8
398	28.5	**512**	28.8	**367**	28.3	**456**	23.3	**340**	23.6
444	28.0	**443**	28.6	**353**	27.7	**353**	23.2	**458**	23.5
297	27.9	**427**	28.4	**428**	26.2	**239**	22.3	**249**	23.3

Table S5.7. Cont.

GaBio		Ara		GalA		GOS-F		CP-I	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
355	82.3	**327**	63.2	**530**	61.0	**355**	47.8	**355**	35.6
322	67.2	**384**	56.0	**355**	56.9	**395**	43.9	**442**	34.1
336	57.2	**355**	54.1	**382**	50.4	**530**	43.7	**382**	33.6
382	51.7	**354**	52.8	**384**	49.5	**382**	41.4	**264**	30.2
442	48.8	**382**	52.5	**383**	49.0	**411**	40.4	**327**	29.9
514	43.2	**165**	48.6	**312**	46.6	**428**	34.4	**383**	29.6
311	43.0	**444**	43.3	**442**	46.4	**327**	33.7	**385**	29.1
383	41.3	**383**	39.4	**327**	43.1	**468**	33.5	**239**	27.7
528	39.5	**442**	38.4	**310**	40.3	**264**	33.5	**428**	27.7
428	38.7	**528**	36.8	**514**	38.7	**528**	32.9	**310**	27.5
385	37.9	**542**	35.6	**354**	38.4	**239**	32.4	**384**	26.5
354	35.7	**514**	34.9	**385**	38.1	**427**	32.3	**530**	26.4
339	34.6	**385**	34.8	**515**	38.0	**512**	31.8	**512**	26.2
165	34.1	**411**	34.7	**311**	37.2	**514**	31.5	**515**	26.1
426	33.4	**264**	34.1	**428**	37.1	**384**	31.4	**528**	25.9
179	32.8	**266**	33.5	**411**	36.1	**385**	30.8	**514**	25.3
411	32.7	**339**	33.2	**165**	35.4	**310**	30.8	**266**	24.7
444	32.2	**267**	33.0	**339**	34.1	**515**	30.5	**324**	23.9
529	32.0	**311**	32.3	**253**	33.1	**339**	30.2	**398**	23.5
512	30.6	**179**	31.7	**251**	32.2	**529**	29.4	**499**	23.3
443	30.3	**409**	30.4	**264**	31.7	**383**	29.0	**294**	23.3
264	29.2	**296**	29.6	**239**	31.6	**354**	28.6	**311**	23.2
327	28.1	**294**	29.6	**179**	29.9	**165**	28.0	**529**	23.1
238	27.7	**456**	27.1	**444**	29.6	**251**	27.6	**354**	23.1
239	26.1	**458**	27.0	**529**	29.3	**253**	27.0	**282**	23.0
515	25.8	**312**	26.9	**249**	28.7	**543**	26.9	**325**	23.0
312	25.6	**323**	25.9	**208**	27.4	**366**	26.6	**542**	22.5
398	24.2	**281**	25.6	**367**	26.7	**398**	26.2	**367**	22.2
297	23.8	**367**	25.6	**443**	26.3	**397**	26.0	**237**	22.1
497	23.7	**239**	25.4	**427**	26.3	**249**	25.8	**543**	22.0

Table S5.7. Cont.

CP-F		AP-F		CPOS-I		CPOS-F		APOS-I	
m/z	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**	***m/z***	**Importance**
530	47.9	**530**	50.1	**530**	71.7	**514**	41.8	**530**	100.0
355	45.5	**514**	48.9	**514**	56.0	**512**	35.4	**442**	55.8
442	44.2	**355**	47.9	**355**	54.7	**442**	34.5	**528**	55.1
382	39.7	**382**	46.3	**442**	52.8	**382**	32.1	**426**	48.1
384	38.6	**442**	45.1	**512**	45.8	**528**	29.8	**296**	42.7
327	38.2	**384**	40.9	**398**	42.1	**426**	29.3	**327**	40.2
339	38.2	**264**	38.2	**426**	42.1	**310**	28.7	**384**	38.9
310	38.1	**366**	37.8	**264**	40.3	**384**	28.0	**339**	38.2
383	37.8	**383**	36.0	**310**	40.1	**264**	27.8	**398**	36.0
385	36.5	**428**	34.9	**528**	39.8	**428**	27.4	**324**	33.0
411	35.1	**512**	34.5	**382**	38.5	**355**	27.1	**294**	33.0
428	34.6	**385**	34.3	**383**	37.8	**515**	26.9	**336**	30.5
251	32.7	**327**	34.1	**165**	37.5	**296**	26.6	**323**	29.1
264	32.0	**310**	34.1	**411**	37.0	**294**	26.1	**427**	29.1
528	31.6	**515**	33.0	**339**	36.4	**511**	25.9	**322**	28.9
514	31.0	**251**	33.0	**428**	35.8	**339**	25.8	**515**	28.8
515	30.8	**528**	33.0	**336**	34.4	**411**	25.7	**542**	28.3
512	30.0	**411**	32.5	**370**	33.4	**336**	25.4	**514**	28.3
165	29.9	**354**	32.5	**311**	32.9	**398**	24.3	**312**	27.7
296	29.8	**311**	31.2	**511**	32.9	**397**	23.8	**382**	27.2
239	29.4	**529**	30.7	**294**	32.7	**309**	23.1	**428**	26.9
529	29.1	**165**	29.8	**441**	32.6	**322**	22.7	**540**	26.7
354	28.9	**511**	28.2	**296**	31.3	**324**	22.6	**440**	26.5
311	28.9	**426**	27.7	**324**	31.2	**529**	22.2	**239**	24.5
426	28.2	**427**	27.5	**542**	30.5	**543**	21.8	**543**	23.4
253	27.8	**179**	27.4	**515**	30.2	**441**	21.7	**396**	23.0
367	27.2	**543**	27.1	**312**	29.6	**513**	21.3	**443**	22.0
427	26.8	**324**	26.3	**427**	28.1	**281**	20.8	**368**	21.4
398	26.4	**239**	26.2	**397**	27.8	**383**	20.5	**497**	21.3
397	26.3	**370**	26.1	**543**	27.6	**266**	20.4	**165**	21.0

Table S5.7. Cont.

APOS-F	
m/z	**Importance**
442	55.3
528	42.5
530	41.5
398	36.4
239	36.0
426	34.5
543	34.4
382	33.7
177	33.6
355	33.2
340	31.4
443	31.3
266	30.2
514	29.5
353	28.7
444	28.2
456	27.2
324	26.9
336	26.4
337	25.8
296	25.3
383	24.7
427	24.0
311	22.7
368	22.0
222	21.7
265	21.6
367	21.6
458	21.4
339	21.1

Table S5.8. Most specific chemical structures of relevant *m/z* ions found in control and functional yogurts used to carbohydrate classification. These ions correspond to TMSO fragments and were determined by CFM-ID.

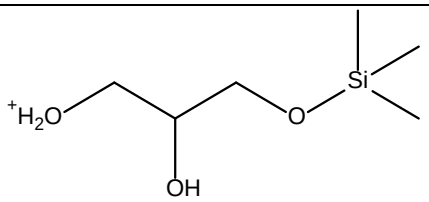
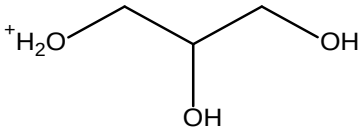
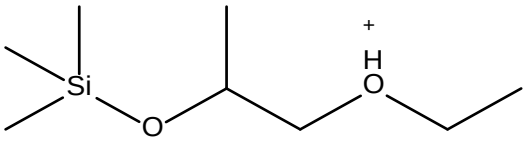
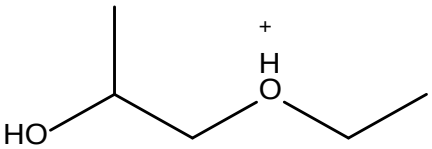
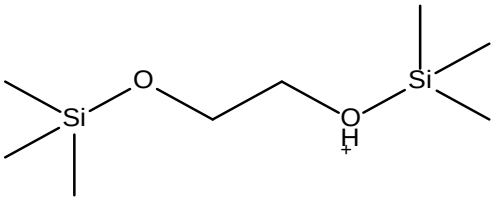
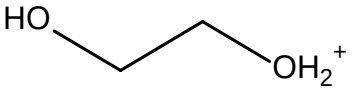
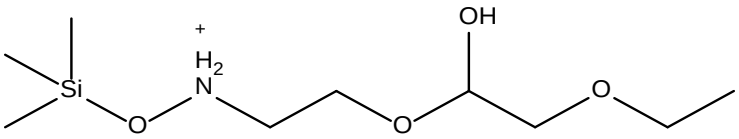
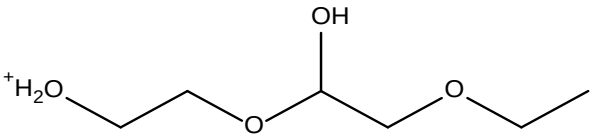
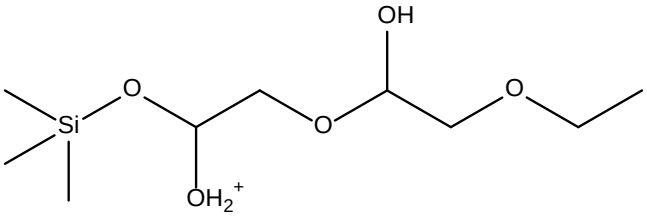
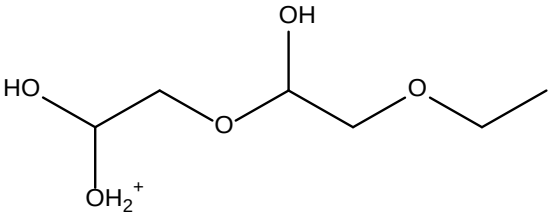
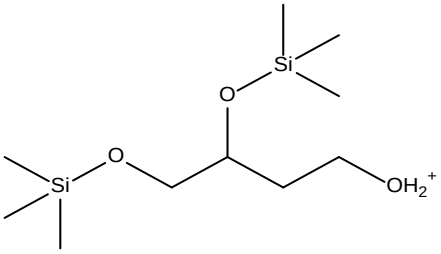
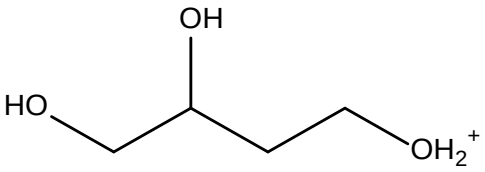
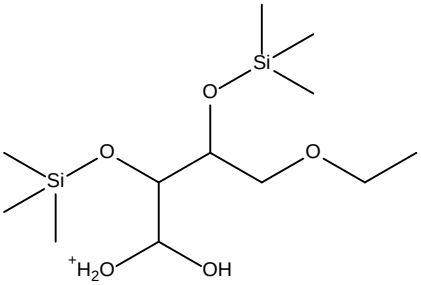
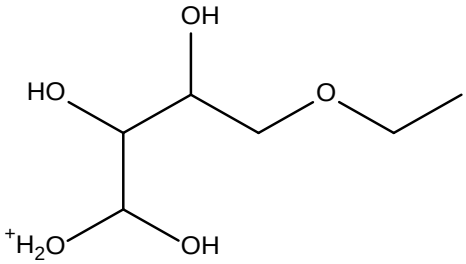
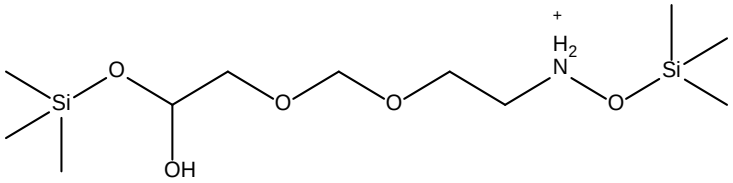
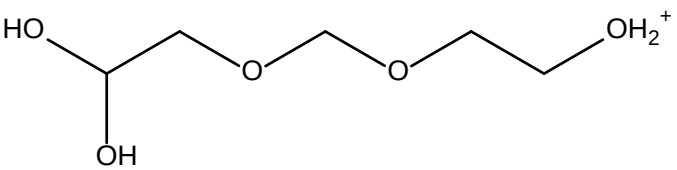
m/z	TMSO structure	Original carbohydrate structure
165		
177		
207		
238		
239		
251		
311		
312		

Table S5.8. Cont.

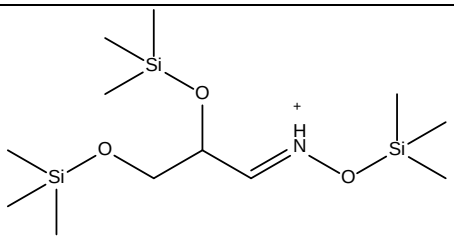
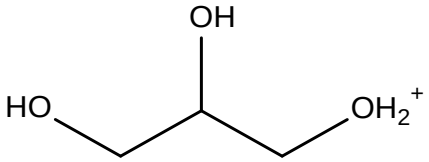
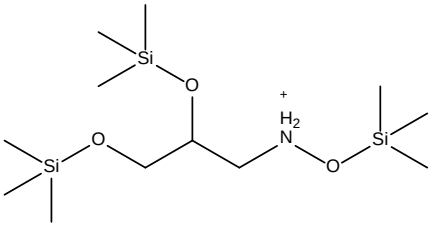
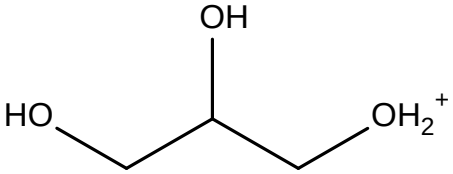
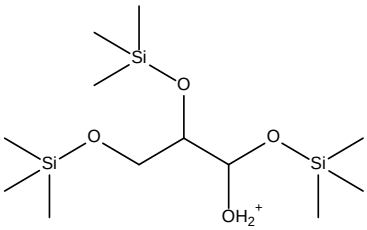
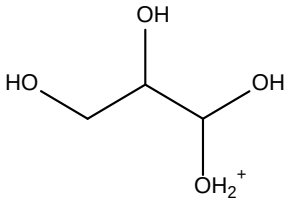
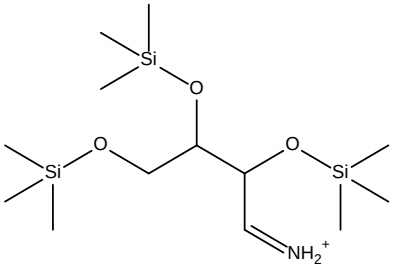
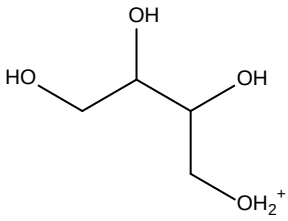
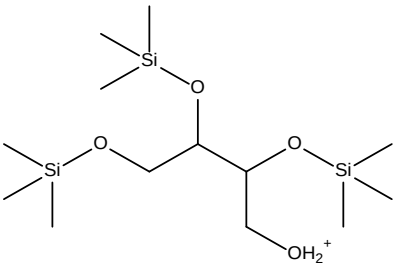
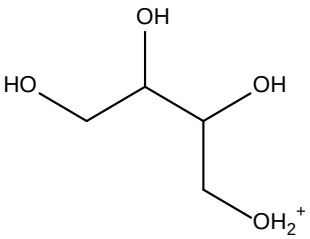
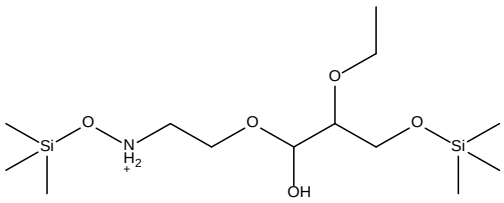
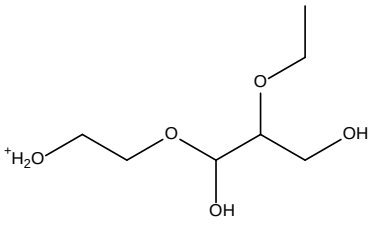
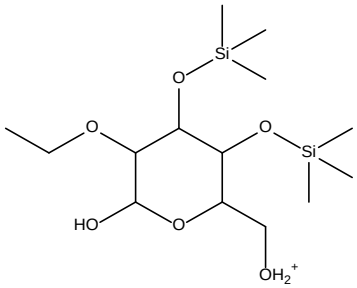
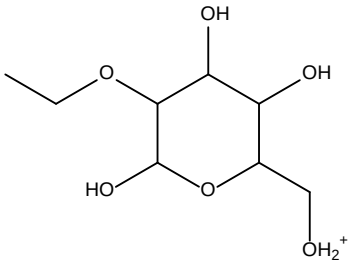
m/z	TMSO structure	Original carbohydrate structure
322		
324		
325		
336		
339		
340		
353		

Table S5.8. Cont.

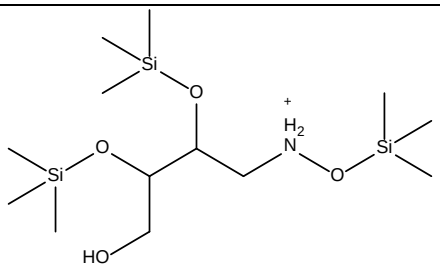
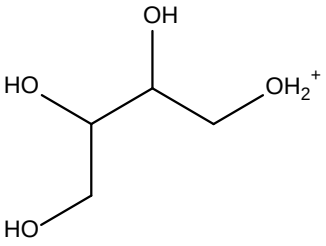
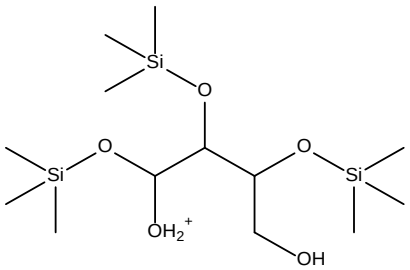
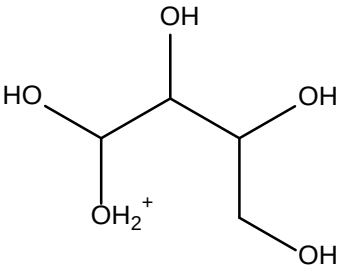
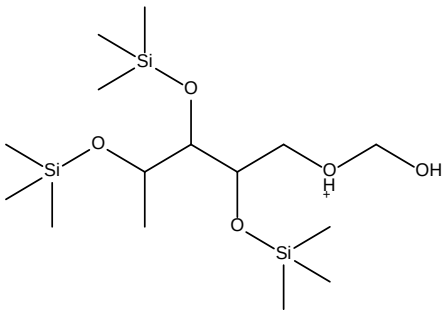
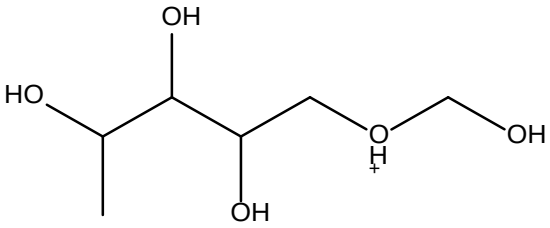
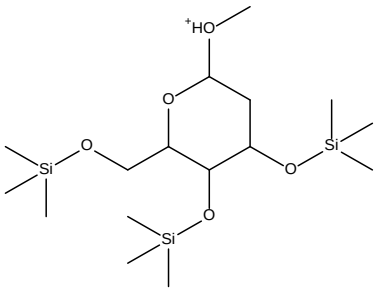
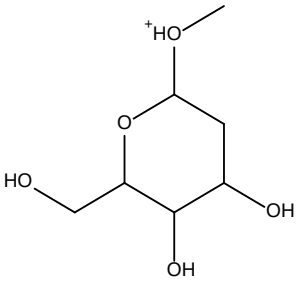
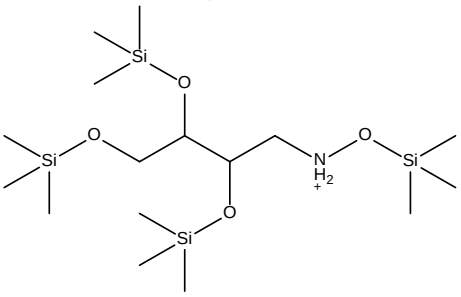
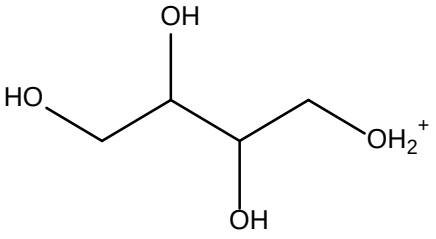
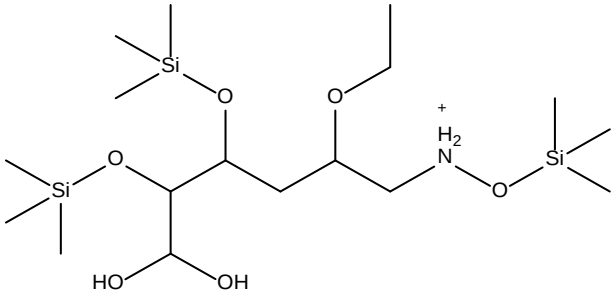
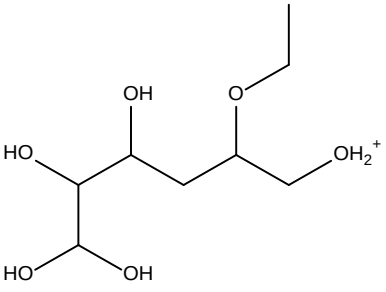
m/z	TMSO structure	Original carbohydrate structure
354		
355		
383		
395		
426		
442		

Table S5.8. Cont.

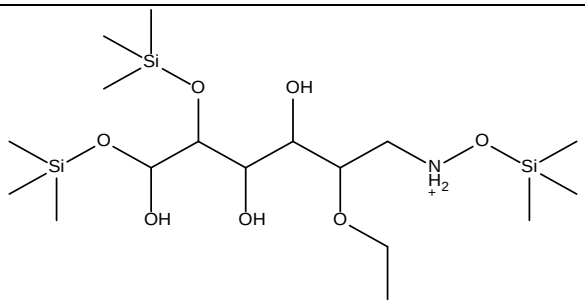
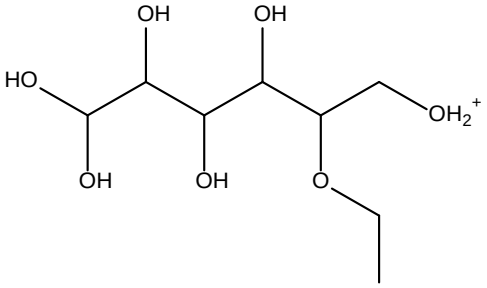
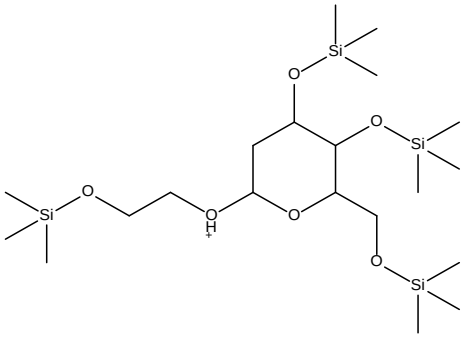
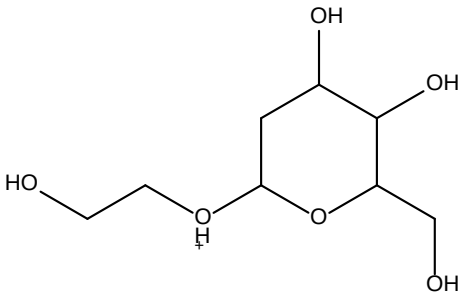
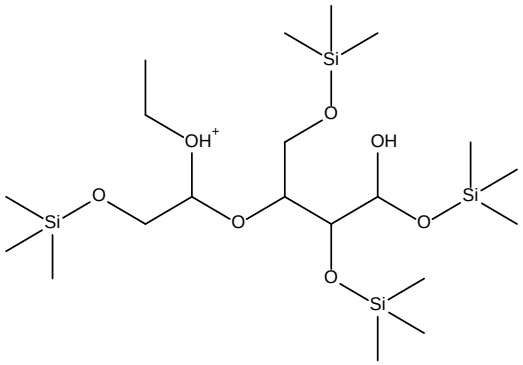
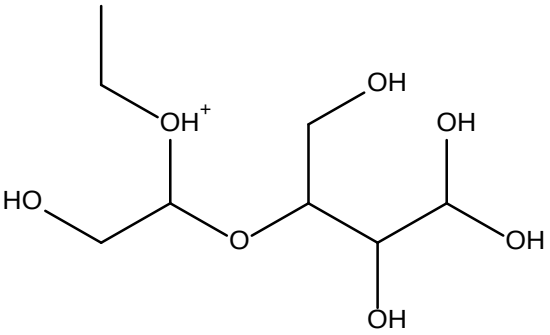
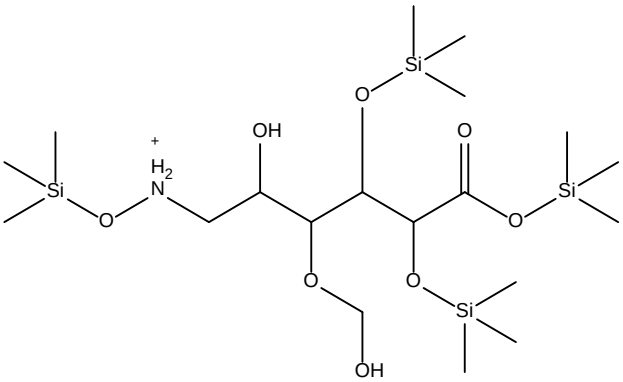
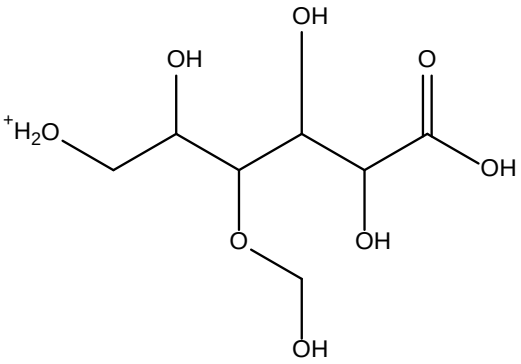
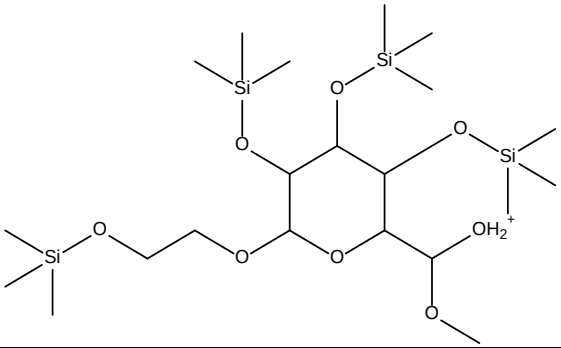
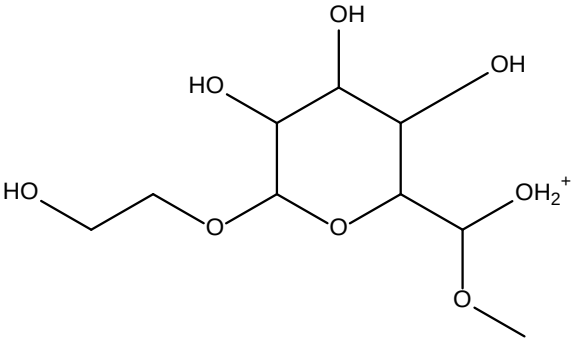
m/z	TMSO structure	Original carbohydrate structure
458		
497		
515		
530		
543		

Table S5.9. Average relative retention times (**RRT**) of the different classes of carbohydrates found in functional yogurts. ***Myo*-In**: myo-inositol, **Gal**: galactose, **Glc**: glucose, **La**: lactose, **AloLa**: allolactose, **Gabio**: 6-galactobiose, **Ara**: arabinose, **GalA**: galacturonic acid, **GOS**: galacto-oligosaccharides, **POS**: pectic oligosaccharides, **CP**: yogurts with citrus pectin, **AP**: yogurts with artichoke pectin, **CPOS**: yogurts with citrus POS, **APOS**: yogurts with artichoke POS, **-I**: initial oligosaccharides, **-F**: formed oligosaccharides.

Class of carbohydrates	RRT*
Ara**	0.309 ± 0.002
Gal	0.704 ± 0.004
Glc	0.719 ± 0.002
GalA	0.779 ± 0.014
***Myo*-In**	0.803 ± 0.001
La	1.109 ± 0.001
AloLa	1.345 ± 0.008
Gabio	1.352 ± 0.010
GOS-F***	1.344 ± 0.033
CP-I	1.201 ± 0.001
CP-F	1.503 ± 0.206
AP-F	1.262 ± 0.076
CPOS-I	1.379 ± 0.205
CPOS-F	1.651 ± 0.001
APOS-I	1.529 ± 0.186
APOS-F	1.486 ± 0.276

* Relative retention times were calculated considering the retention time of each molecule and internal standard (β-phenil glucoside).

** Two peaks were considered for reducing sugars where the exception of galactose and glucose, where the second peaks coeluted.

*** Different oligosaccharide peaks including from di- to tetrasaccharides were considered for GOS and POS. This table shows the average RRT values for each class.

Table S5.10. Root mean squared error (**RMSE**), $\mathbf{R^2}$ and mean adjustment error (**MAE**) and their adjusted values, obtained for each regression model: bagged classification and regression trees (**BCART**) and random forest (**RF**).

Train phase				
Model	**RMSE**	$\mathbf{R^2}$	$\mathbf{R^2_{adj}}$	**MAE**
BCART	0.148	0.883	0.874	0.112
RF	0.016	0.999	0.998	0.010
QRF	0.001	0.999	0.999	0.001
Test phase				
Model	**RMSE**	$\mathbf{R^2}$	$\mathbf{R^2_{adj}}$	**MAE**
BCART	0.153	0.860	0.832	0.110
RF	0.036	0.992	0.991	0.022
QRF	0.005	0.999	0.999	0.001

Table S5.11. Most influential *m/z* ions in quantitative structure-relative retention time relationships using random forest (**RF**) and quantile regression forest (**QRF**) models. For QRF importance coefficients were determined by calculating the area under the ROC (Receiver Operating Characteristic) curve while importance coefficients were determined by the permutation of the out of-bag-error in RF.

	Importance	
m/z	**RF**	**QRF**
165	24.1	46.5
177	28.8	27.0
207	15.1	35.3
238	26.8	33.2
239	22.0	52.7
251	29.1	42.1
311	15.3	20.8
312	18.9	34.6
322	60.6	89.0
324	75.2	100.0
325	11.7	15.0
336	16.5	16.8
339	15.9	6.9
340	46.1	72.3
353	21.2	23.3
354	49.7	66.6
355	21.5	36.1
383	19.1	28.7
395	16.0	35.4
426	19.9	3.0
442	16.4	41.0
458	10.2	4.8
497	14.8	18.6
515	23.8	51.7
530	21.5	31.0
543	11.7	0.0

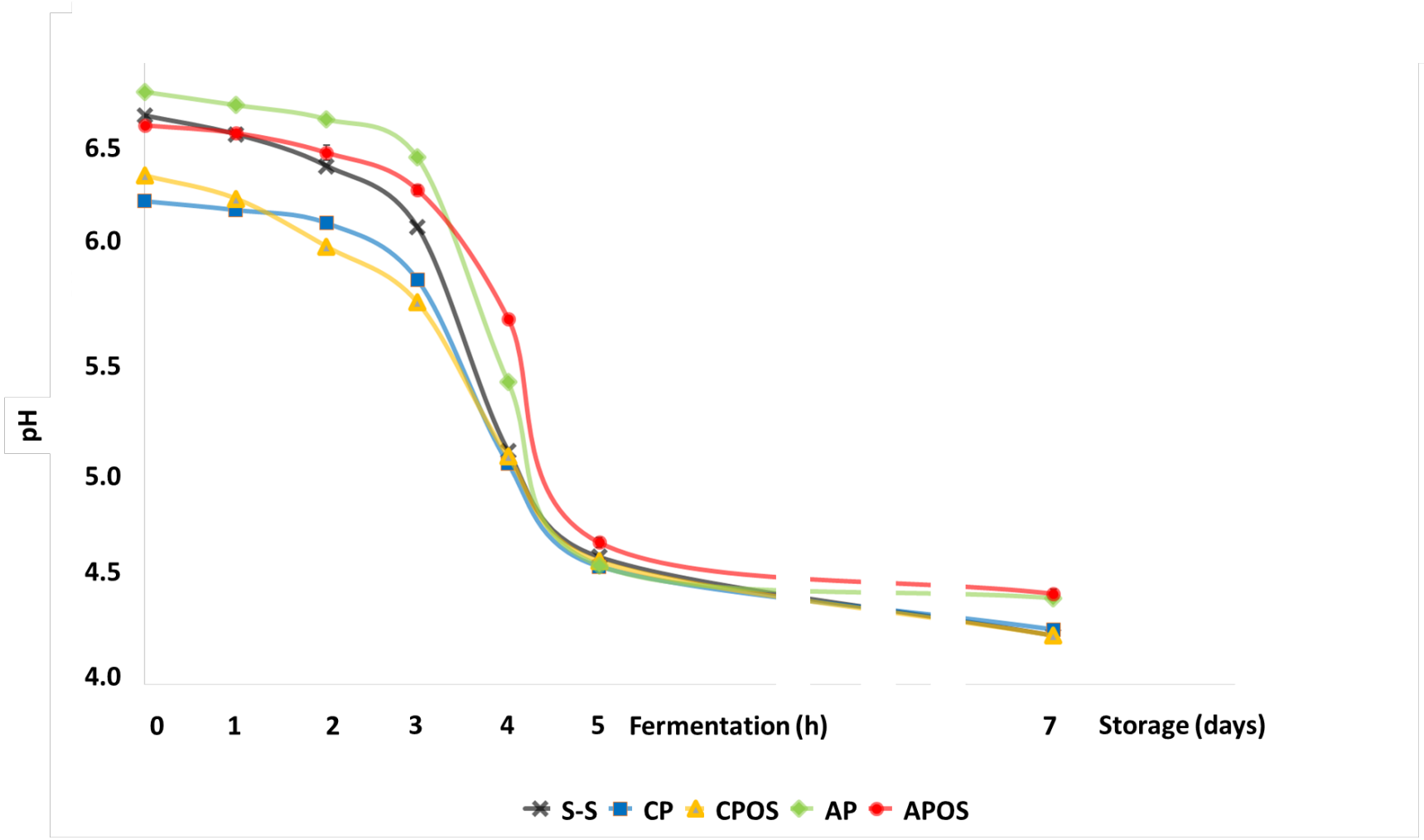


Figure S5.1. Monitorization of pH during fermentation and cold storage of set-style (**S-S**) yogurt and yogurts containing citrus pectin (**CP**), citrus POS (**CPOS**), artichoke pectin (**AP**) and artichoke POS (**APOS**).

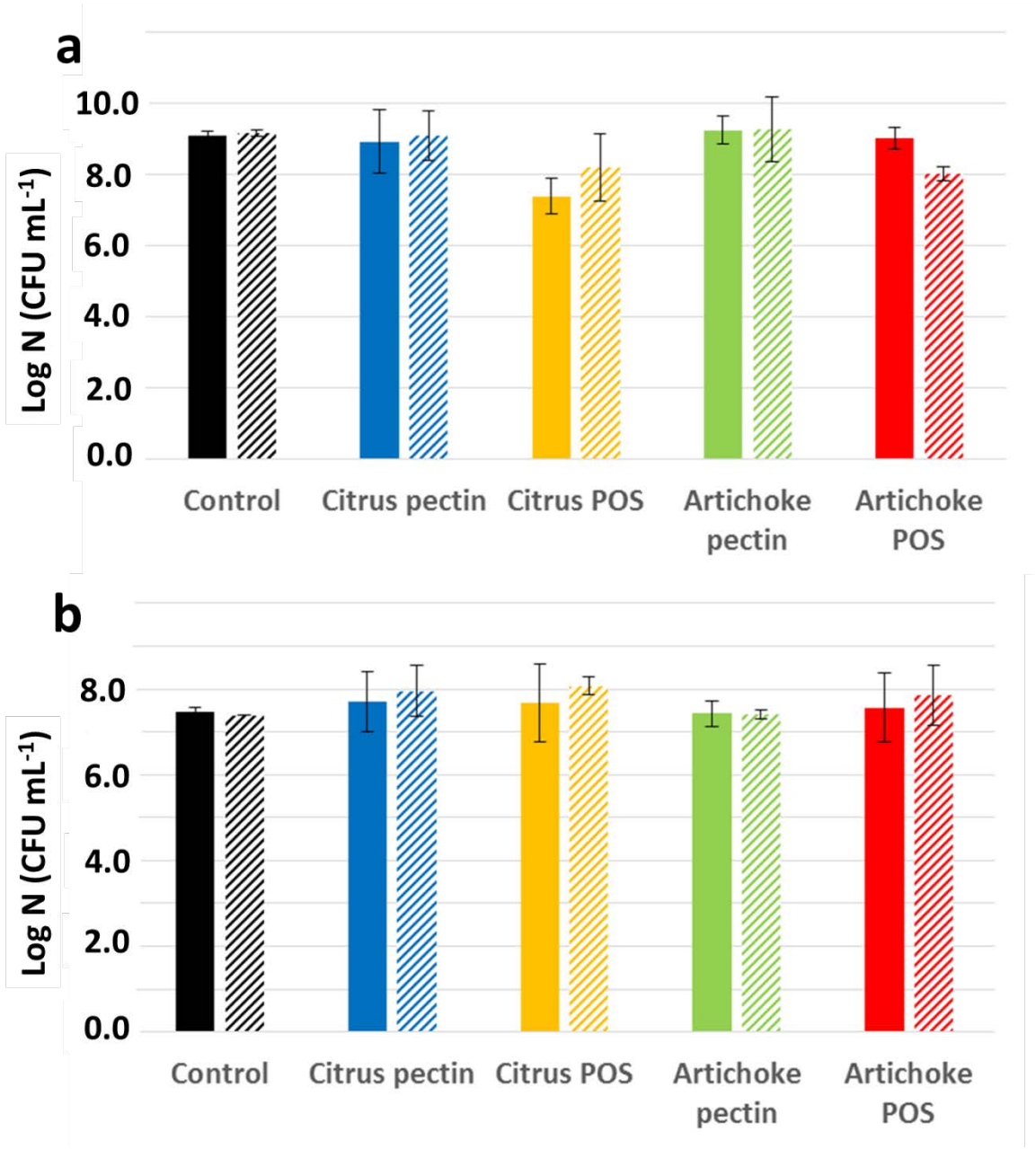


Figure S5.2.Viability of *S. termophilus* (**a**) and *L. bulgaricus* (**b**) cell count after fermentation (solid) and 1 week of cold storage (frame) in set-style and functional yogurts. No statistically significant differences were found between groups.

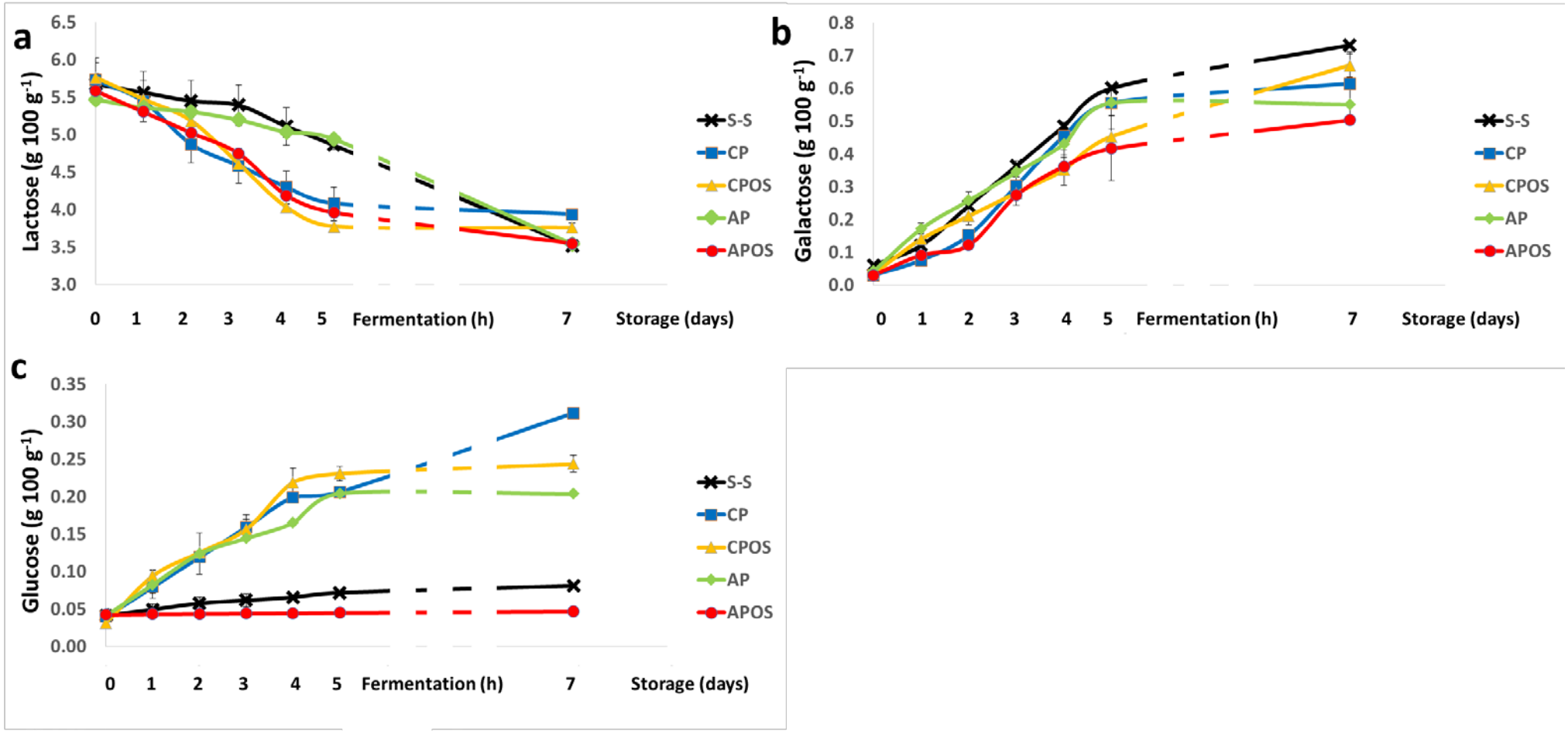


Figure S5.3. Evolution of lactose (**a**), galactose (**b**), glucose (**c**) during fermentation and cold storage in set-style (**S-S**) yogurt and yogurts containing citrus pectin (**CP**), citrus POS (**CPOS**), artichoke pectin (**AP**) and artichoke POS (**APOS**).

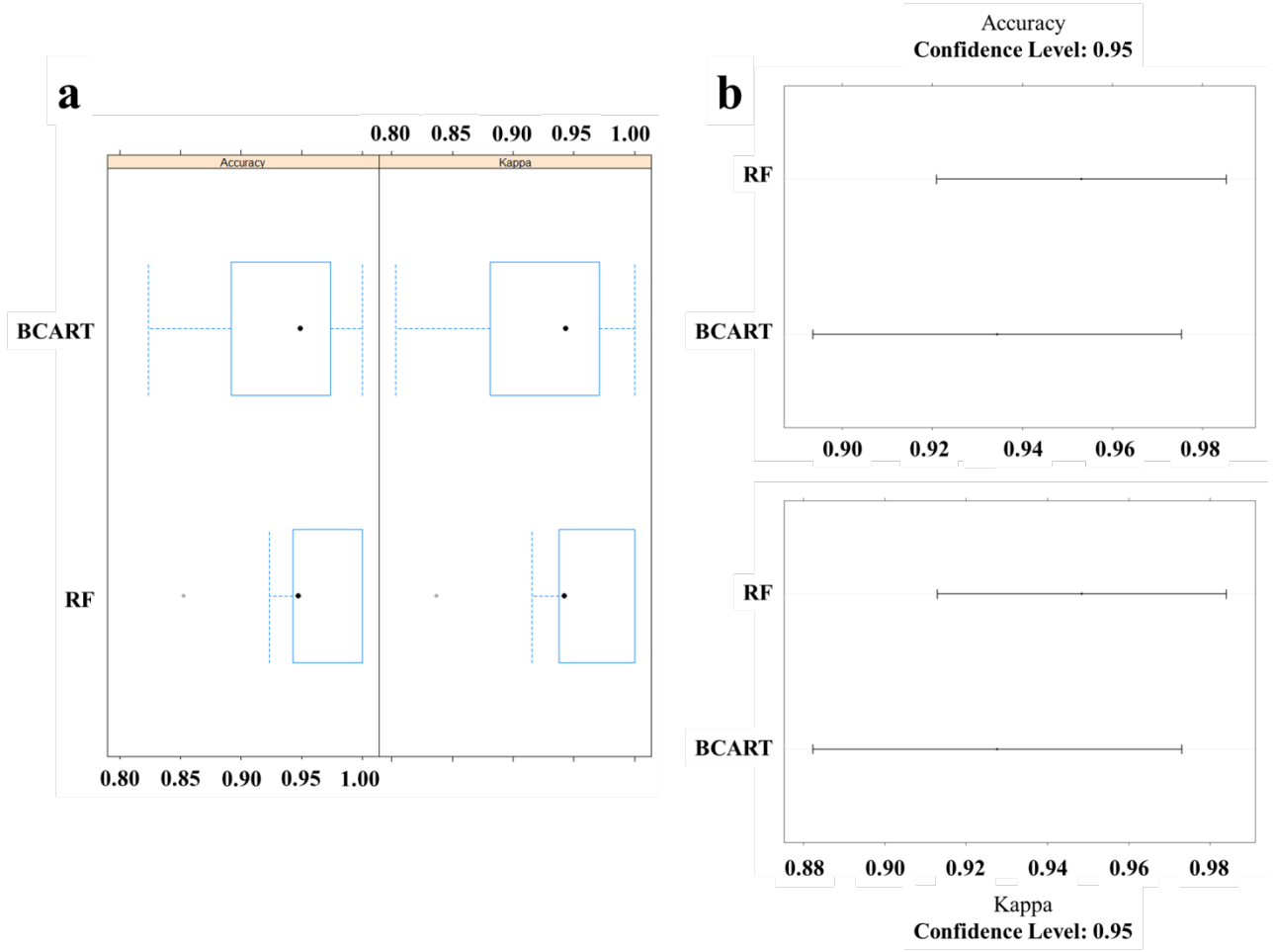


Figure S5.4. Comparative account of bagged classification and regression trees (**BCART**) and random forest (**RF**). **a)** boxplots of accuracy and kappa values for each models, **b)** accuracy and kappa values obtained at a confidence level of 0.95. Differences between models were calculated via their resampling distributions (number of resamples = 10). No statistically significant differences between algorithms were found.

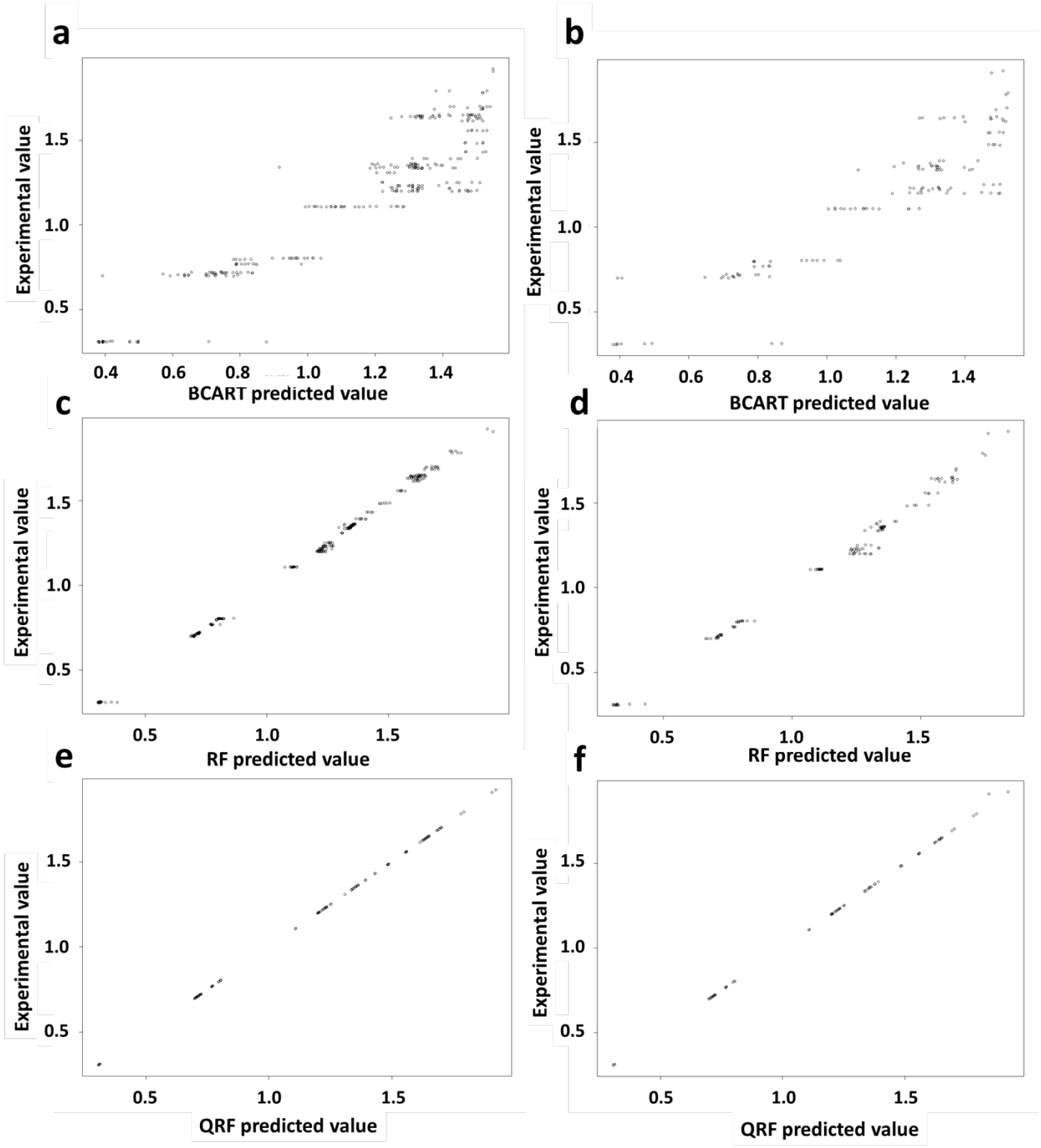


Figure S5.5. Experimental and predicted RRT values: **a)** BCART (train phase), **b)** BCART (test phase), **c)** RF (train phase), **d)** RF (test phase), **e)** QRF (train phase), **f)** QRF (test phase).

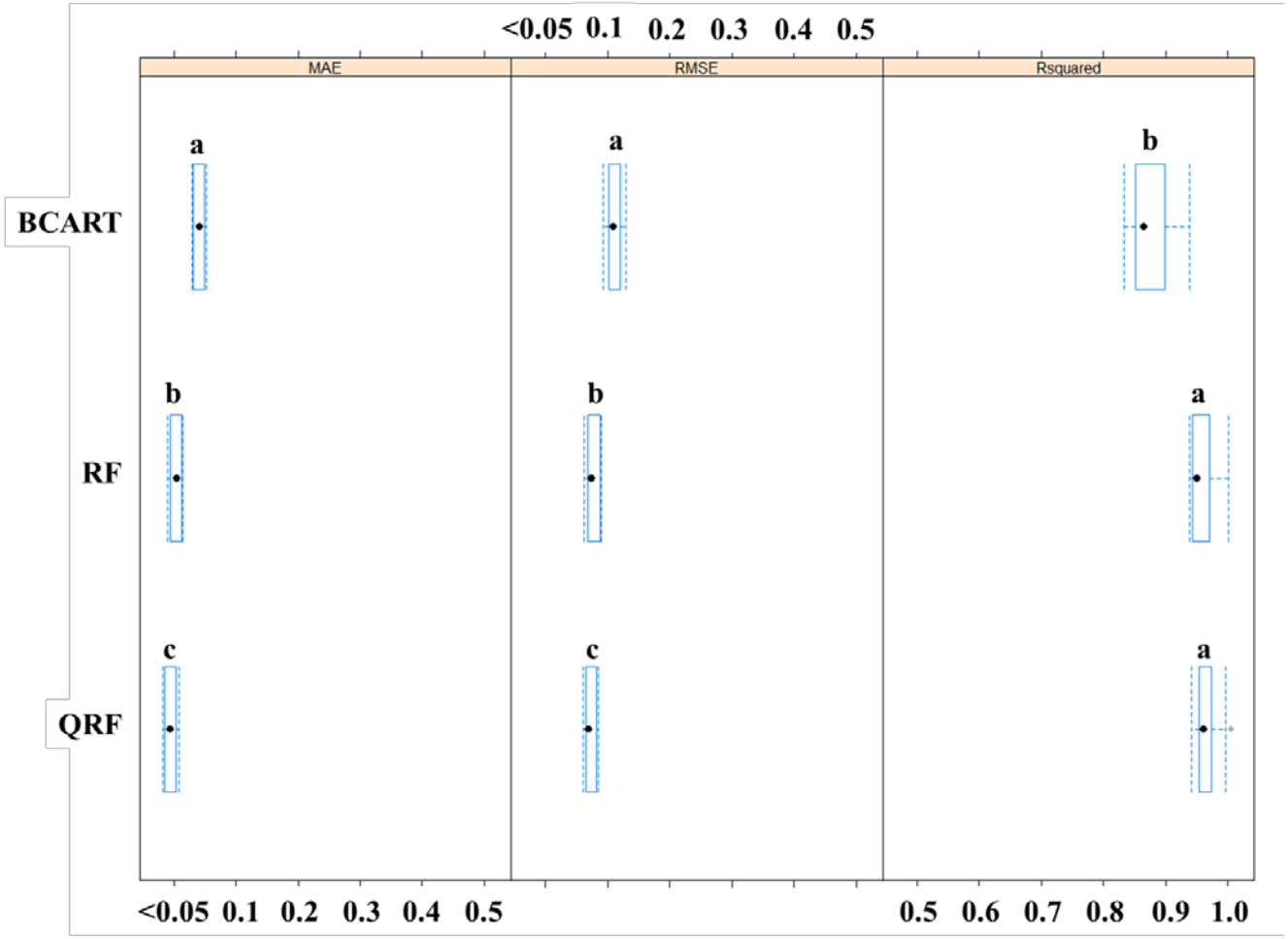


Figure S5.6. Comparative account of bagged classification and regression trees (**BCART**), random forest (**RF**) and quantile regression forest (**QRF**) showing differences in Root mean squared error (**RMSE**), mean adjustment error (**MAE**) and R^2 calculated via their resampling distributions (number of resamples = 10). [a, b] Statistically significant differences between models

Annex I.VI:

Supplementary material of Article VI

11.6. Annex I.VI: Supplementary material of Article VI

Table S6.1. Clinical parameters considered to determine the Disease Activity Index (DAI) in healthy, DSS treated and DSS + pectins treated mice groups.

Bleeding	**Stool consistency**	**Weight loss (WL, %)**	**Value assigned according to WL**
0: normal	**0**: normal	0	0
1: presence of blood	**1**: moderate soft stools	1 - 4	1
2: moderate bleeding	**2**: soft stools	5 – 9	2
3: moderately high bleeding	**3**: soft stools and diarrhea	10 – 19	3
4: abundant bleeding	**4**: diarrhea	> 20	4

Table S6.2. Monomeric composition of citrus (CP) and artichoke pectin (AP) and modified artichoke pectin fractions used in this study: APwA: modified artichoke pectin without arabinose, APwG: modified artichoke pectin without galactose, GalA: galacturonic acid, Xyl: xylose, Ara: arabinose, Rha: rhamnose, Fru: fructose, Gal: galactose, Man: mannose, Glc: glucose.

Sample	Monomeric composition (% total identified monosaccharides)								Degree of branching Rha/GalA	Linearity pectin backbone GalA/(Rha+Ara+Gal)	Extent of branching (Ara+Gal)Rha	Degree of methyl-esterification (%)
	GalA	Xyl	Ara	Rha	Fru	Gal	Man	Glc				
CP*	75.51^{a,b} (0.54)	2.51^{a} (0.02)	3.18^{c} (0.01)	4.52^{a,b} (0.05)	-	11.10^{a} (0.10)	0.30^{b} (0.00)	0.98^{c} (0.01)	0.060^{a} (0.00)	4.02^{a} (0.01)	3.16^{b} (0.03)	71.0^{a} (1.4)
AP	68.68^{b} (1.23)	0.53^{c} (0.19)	15.92^{a} (0.03)	3.25^{b} (0.18)	0.55^{a} (0.07)	6.38^{b} (0.57)	1.35^{a} (0.09)	3.33^{b} (0.42)	0.047^{a} (0.003)	2.69^{a} (0.13)	6.86^{a} (0.20)	19.5^{b} (0.0)
APwA	76.04^{a,b} (3.45)	0.17^{c} (0.05)	1.38^{c} (0.16)	6.32^{a} (0.93)	<0.10^{b} (0.00)	9.69^{a} (1.35)	0.47^{b} (0.11)	5.93^{a} (0.84)	0.083^{a} (0.016)	4.43^{a} (0.82)	1.75^{d} (0.02)	24.0^{b} (1.9)
APwG	77.26^{a} (1.66)	1.52^{b} (0.22)	12.57^{b} (0.88)	4.93^{a,b} (0.49)	<0.10^{b} (0.00)	0.62^{c} (0.05)	0.30^{b} (0.03)	2.80^{b,c} (0.01)	0.064^{a} (0.008)	4.28^{a} (0.43)	2.68^{c} (0.08)	24.3^{b} (0.9)

*Data obtained from Pacheco et al. (2018)
^{a,b,c,d} Statistically significant differences between pectins.

Table S6.3. Molecular weight (M_w) distribution of citrus (CP) and artichoke pectin (AP) and modified artichoke pectin fractions used in this study. APwA: modified artichoke pectin without arabinose, APwG: modified artichoke pectin without galactose.

	CP	**AP**	**APwA**	**APwG**
M_w of each fragment [abundance, %]	547 ± 11 kDa [100 ± 0%]	660 ± 0 kDa [28.5 ± 0.2%]	542 ± 12 kDa [41.7 ± 1.3%]	605 ± 15 kDa [31.0 ± 0.8%]
		105 ± 3 kDa [36.6 ± 0.1%]	63 ± 3 kDa [58.3 ± 1.3%]	74 ± 8 kDa [39.5 ± 2.4%]
		4.8 ± 1.2 kDa [34.9 ± 0.2%]		5.1 ± 0.5 kDa [29.5 ± 1.6%]

Table S6.4. Train, cross-validation, test rates and kappa values for all artificial neural network (ANN) models developed to study characteristic expression profiles of cytokine and intestinal proteins in studied groups of mice. DSS: dextran sulfate sodium, CP: citrus pectin control, AP: artichoke pectin, APwA: modified artichoke pectin without arabinose, APwG: modified artichoke pectin without galactose.

Parameter	ANN-1 (Healthy control)	DSS treatment				
		ANN-2 (Control)	ANN-3 (+ CP)	ANN-4 (+ AP)	ANN-5 (+ APwA)	ANN-6 (+ APwG)
Train rate (%)	100	100	100	100	100	100
Cross-validation rate (%)	100	100	97.5	95.5	92.6	98.0
Test rate (%)	100	94.7	94.1	94.4	94.4	100
Kappa	1.00	0.94	0.94	0.85	0.82	0.98
Sensitivity	100	100	100	90.0	100	100
Specificity	100	94.1	93.3	100	93.3	100
Balanced accuracy	100	97.1	96.7	95.0	96.7	100

Annex II:

Published articles

Carbohydrate Polymers 190 (2018) 43–49

Contents lists available at ScienceDirect

Carbohydrate Polymers

journal homepage: www.elsevier.com/locate/carbpol

Enzymatic extraction of pectin from artichoke (*Cynara scolymus L.*) by-products using Celluclast®1.5L

Carlos Sabater, Nieves Corzo*, Agustín Olano, Antonia Montilla

Instituto de Investigación en Ciencias de la Alimentación CIAL, (CSIC-UAM) CEI (UAM + CSIC), C/Nicolás Cabrera, 9, E-28049 Madrid, Spain

ARTICLE INFO

Keywords:
Pectin extraction
Artichoke
Cynara scolymus
By-products
Celluclast
RSM

ABSTRACT

The aim of this study was to optimise pectin extraction from artichoke by-products with Celluclast®1.5L using an experimental design analysed by response-surface methodology (RSM). The variables optimised were artichoke by-product powder concentration (2–7%, X_1), enzyme dose (2.2–13.3 U g^{-1}, X_2) and extraction time (6–24 h, X_3). The variables studied were galacturonic acid (GalA) (R^2 93.9) and pectic neutral sugars (R^2 92.8) content and pectin yield (R^2 88.6). In the optimum extraction conditions (X_1 = 6.5%; X_2 = 10.1 U g^{-1}; X_3 = 27.2 h), pectin yield was 176 mgg^{-1} dry matter (DM). Considering 27.2 h of treatment as the +α value given by the design, the extraction time was increased up to 48 h obtaining a yield of 221 mg g^{-1} DM. The enzymatic method optimised allows obtaining artichoke pectin with good yield, high GalA (720 mg g^{-1} DM) and arabinose (127.6mgg^{-1} DM) contents and degree of methylation of 19.5%.

1. Introduction

Pectins are a type of soluble fibre commonly present in vegetal cell wall and are one of the most structurally complex families of polysaccharides in nature. It is mainly composed of a linear chain of α-1,4-D-galacturonic acid (GalA) called homogalacturonan (HG) which comprises approximately 70% of pectin. This polymer is partially methylesterified at the C-6 carboxyl and may be *O*-acetylated at O-2 or O-3. HG is the backbone that presents ramified chains of rhamnogalacturonan type I (RGI), a polymer made up of alternate sequences of GalA and α-(1, 2) linked L-rhamnosyl residues, and rhamnogalacturonan II (RGII), a most complex chain consisting of 12 different types of sugars in over 20 different linkages (Garna, Mabon, Nott, Wathelet, & Paquot, 2006; Gullón et al., 2013; Mohnen, 2008).

Pectins are widely used as valuable functional food ingredients due to their technological and biological properties. The majority of pectins used are extracted from citrus peel, apple and sugar beet by-products (Santos, Espeleta, Branco, & de Assis, 2013). Functional properties, which depend on structure of pectins, are influenced by sources and conditions of extraction, location and other environmental factors (Canteri, Nogueira, de Oliveira Petkowicz, & Wosiack, 2012; Fissore et al., 2014; Gullón et al., 2013). In the last years non-traditional sources of pectins have also been studied such as fresh peas pomace, sunflower heads, bark of mango tree, red fruit pulps and sisal wastes (Santos et al., 2013). In the same way, other new source of pectin could be artichoke by-products. Artichoke (*Cynara scolymus* L.) is an edible vegetable widely consumed in the Mediterranean diet, harvesting in this area the 85% of world production. The industrial processing of artichoke produces a high amount of wastes (external bracts, leaves and stems), which is close to 60% of the harvested artichoke (Llorach, Espin, Tomás-Barberán, & Ferreres, 2002). Although artichoke residues are generally used for animal feed or as manure (Machado et al., 2015), they can be used to extract bioactive compounds such as fructooligosaccharides and inulin (López-Molina et al., 2005; Schütz, Muks, Carle, & Schieber, 2006; Terkmane, Krea, & Moulai-Mostefa, 2016); flavonoids and phenolic compounds (Llorach et al., 2002; Sałata & Gruszecki, 2010) and inositols (Ruiz-Aceituno, García-Sarrió, Alonso-Rodriguez, Ramos, & Sanz, 2016) as well as to isolate fractions enriched in soluble fibre (Fissore et al., 2014). However, there are scarce studies about extraction and characterisation of pectin from artichoke by-products (Femenia, Robertson, Waldron, & Selvendran, 1998; Fishman, El-Atawy, Sondey, Gillespie, & Hicks, 1991; Fissore et al., 2014).

The procedures for the extraction of commercial pectins are usually optimised to enhance the content of HG regions which generate the useful functional gelling properties of pectin (Maxwell, Belshaw, Waldron, & Morris, 2012). With this finality, the chemical extraction of pectins, employing strong acids such as hydrochloric, nitric, and sulphuric acids and high temperature (70–100 °C), has been the most widely used method for the industry (Min et al., 2011; Yapo, 2009) and although is an efficient and economical procedure, may be environmentally hazardous and could modify the structure of pectin.

Extraction of pectins can also be performed through the use of

* Corresponding author.
E-mail addresses: carlos.sabater@csic.es (C. Sabater), nieves.corzo@csic.es (N. Corzo), a.olano@csic.es (A. Olano), a.montilla@csic.es (A. Montilla).

https://doi.org/10.1016/j.carbpol.2018.02.055
Received 20 November 2017; Received in revised form 30 January 2018; Accepted 19 February 2018
Available online 21 February 2018

enzymes and specifically cellulases and xylanases which give rise to pectins without structural modifications. The type of enzyme could determine the yield of extraction and the composition of pectin (Dominiak et al., 2014; Wikiera, Mika, Starzyńska-Janiszewska, & Stodolak, 2015a). Among commercial enzymatic preparations with cellulase activity, Celluclast®1.5L (Celluclast) led to complete pectin extraction with hardly any modification in its structure; thus, this enzyme has been used to extract pectin from apple pomace due to its xylano-, cellulo-, and mannolytic activities (Wikiera et al., 2015a). In addition, pectin extracted with Celluclast contained more GalA of higher methylation and acetylation degree compared to pectin extracted with acids (Wikiera, Mika, Starzyńska-Janiszewska, & Stodolak, 2015b). Therefore, the aim of this study was to optimize pectin extraction from by-products (external bracts, leaves and stems) of artichoke *(Cynara scolymus)* industrialisation using the commercial enzymatic preparation Celluclast. For this purpose, a three variable central composite orthogonal design analysed by response-surface methodology (RSM) was developed.

2. Materials and methods

2.1. Standards and samples

Analytical reference substances such as sucrose, D-arabinose, L-rhamnose, D-galactose, D-mannose, D-glucose, D-fructose, galacturonic acid (GalA), and β-phenyl glucoside were purchased from Sigma Aldrich (Steinheim, Germany). Eight commercial multienzymatic preparations were studied. Celluclast®1.5L (cellulase from *Trichoderma reesei*), Pectinex Ultra SP-L® (pectinase from *Aspergillus aculeatus*) Viscozyme® L (*endo*-1,3(4)-β-glucanase from *A. aculeatus*), Viscozyme® Barley HT (amylase and cellulase from *Aspergillus* sp.) and Carezyme® 4500 L (cellulase from *Aspergillus* sp.) were generous gifts from Novozymes (Bagsvaerd, Denmark). Pectinase and cellulase from *A. niger* and cellulase from *T. viride* were acquired from Sigma.

Cellulase and pectinase activities were measured by quantifying the amount of reducing sugars released, using the dinitrosalicylic acid (DNS) method (Miller, 1959). To determine cellulase activity, cellulose solutions (10 mg mL^{-1}, Carboxymethyl cellulose, Sigma) dissolved in sodium acetate buffer (0.05 M, pH 5) with 5 µL mL^{-1} or 1 mg mL^{-1} of enzyme at 50 °C for 5 min were prepared and glucose was employed as standard. One unit of cellulase activity was defined as the amount of the enzyme, which released 1 µmol of glucose per min. Pectinase activity was measured using polygalacturonic acid solutions (10 mg mL^{-1}, Sigma) incubated with 5 µL mL^{-1} or 1 mg mL^{-1} of enzyme at 50 °C for 2 min. GalA was employed as a standard. One unit of pectinase activity was defined as the amount of the enzyme, which released 1 µmol of GalA per min.

Industrial by-product of artichoke, consisting in blanched external bracts, leaves and stems, were kindly supplied by Riberebro® (La Rioja, Spain). Prior to characterisation, this product was freeze-dried, homogenised using a Wiley Mill 934 equipment (Thomas Scientific, USA) and sieving through a 500 µm sieve (Femenia et al., 1998).

2.2. Pectin extraction from artichoke by-products

Enzymatic extraction of pectin from artichoke by-products (particle size ≤ 500 µm) was performed using Celluclast. The process was conducted in an orbital shaker at 50 °C, pH 5 with constant shaking (200 rpm). After hydrolysis, samples were centrifuged (1300*g*, 10 min, 4 °C) and supernatants were filtered through cellulose paper. The resultant precipitate was washed with distilled water (20 mL), centrifuged and filtered. The supernatants were combined and ethanol at 4 °C was added to obtain a final concentration of 70%. After one hour at 4 °C, the precipitated pectins were centrifuged (1200*g*, 20 min), washed twice with 70% ethanol (600 mL), centrifuged again, and freeze-dried.

2.2.1. Experimental design for extraction of pectins

A three level, three variable, central composite orthogonal design was applied to determine the best conditions to extract enzymatically pectin from artichoke by-products. The factorial design consisted of 8 factorial points, 6 axial points (two axial points on the axis of each design variable at a distance of 2 from the design centre) and 3 centre points. The independent variables used in this study were artichoke by-product powder concentration (2–7% w/w, X_1), enzyme dose (2.2–13.3 U g^{-1}, X_2), and extraction time (6–24 h; X_3). Four dependent variables were taken into account to optimise pectin extraction by means of RSM: yield of pectin extraction (mg of extracted pectin/g dry matter (DM), Y_1), GalA content (mg g^{-1} DM, Y_2), pectic neutral sugars content (arabinose, rhamnose and galactose, mg g^{-1} DM, Y_3) and other neutral sugars content (glucose, mannose and fructose, mg g^{-1} DM, Y_4).

The responses obtained from each set of experimental design were subjected to multiple nonlinear regressions to achieve the coefficients of the second order polynomial model. The quality of the fit of the polynomial model equation was expressed by the coefficient of determination R-squared (R^2) and the adjusted R-squared (R^2adj), which provides a measurement of how much of the variability in the observed response values could be explained by the experimental factors and their interactions (Myers, Montgomery, Vining, Borror, & Kowalski, 2004). Variables that presented R^2adj values higher than 70% were submitted to simultaneous optimisation.

The overall effect of the three independent factors was used to obtain a desirability function that represents the influence of the extraction conditions on the efficiency of enzymatic extraction with Celluclast and on the final product composition. In general terms, a desirability function represents the effect of the independent variables (X_i) on the dependent variables (Y_j), determining the process efficiency (Myers et al., 2004). For each extraction conditions, the desirability function assigns numbers between 0 (completely undesirable value) and 1 (completely desirable or ideal response). The method allows optimize extraction conditions that provide the “most desirable” response values.

Additional extractions of artichoke pectin were carried out at following conditions: artichoke by-product powder 6.5% (w/w), enzyme dose 10.1 U g^{-1} and extraction time 36 and 48 h.

2.3. Characterisation of samples

All analyses were carried out in duplicate.

2.3.1. Artichoke by-products and extracted pectin

Dry matter (DM) was determined gravimetrically drying the samples until constant weight according to the AOAC method 950.01 (Association of Official Analytical Chemists (AOAC), 1990a).

pH was measured in solutions of 10% using a Mettler Toledo Five Easy Plus pH Meter.

Water activity (a_w) measurement was carried out in an AW Sprint TH-500 instrument (Novasina, Pfäffikon, Switzerland). Saturated aqueous solutions of LiCl, $MgCl_2$, $Mg(NO_3)_2$, NaCl, $BaCl_2$ and $K_2Cr_2O_7$ were used to calibrate the sensor unit.

Protein was determined following the Kjeldhal method for total nitrogen (TN) using a conversion factor (TN × 6.25) to express results as protein (Association of Official Analytical Chemists (AOAC), 1990b).

2.3.2. Artichoke by-product powder

Fat content was measured following AOAC 920.39 (Association of Official Analytical Chemists (AOAC), 1990c) methodology performing Soxhlet extraction using a Gerhardt soxtherm extractor (Germany).

Water soluble (WSF) and insoluble (WIF) fractions were obtained by immersing 1.5g of powdered samples in 40mL boiling water for 30min. The mixture was cooled and filtered through a sintered glass filter (No 2). The residue was re-extracted twice with boiling water and then washed with 40 mL of water. Supernatant was collected and then

residue was washed with 40 mL of acetone 70% and 40 mL of ethanol. Water insoluble fraction (WIF) was dried and stored for further analysis. Supernatant (water soluble fraction, WSF) was partially dried at 38–40 °C in a rotary evaporator (Büchi Labortechnik AG, Flawil, Switzerland), freeze-dried, and redissolved in water at 10 mg mL^{-1} for chromatographic analysis.

Alcohol soluble (ASF) and insoluble (AIF) fractions were obtained by immersing 1.5 g of powdered samples in 40 mL boiling ethanol, final concentration 80% (v/v), for 30 min. The mixture was cooled and filtered through a sintered glass filter (No 2). The residue was re-extracted twice at the same conditions. Supernatants were collected and then residue was washed first with 40 mL of 70% acetone and then with 40 mL of ethanol. Alcohol insoluble fraction (AIF) was dried and stored for further analysis. Supernatant (alcohol soluble fraction, ASF) was partially dried at 38–40 °C in a rotary evaporator, freeze-dried and redissolved in water at 10 mg mL^{-1} for chromatographic analysis.

2.4. Analytical techniques

2.4.1. ICP-MS

The ion composition of artichoke by-product powder and extracted pectin was determined using an ICP-MS ELAN 6000 Perkin-Elmer Sciex instrument at the Servicio Interdepartamental de Investigación (SIdI-UAM) of Madrid. Quantitative analysis of the elements of interest using the external calibration method and internal standard to correct instrumental drift was carried out.

2.4.2. FT-IR

In order to determine the degree of methylation, freeze-dried samples of artichoke by-products and extracted pectin were analysed by FT-IR analysis. KBr discs were prepared mixing the pectin samples with KBr (1:100) and pressed. FT-IR spectra were performed in a Bruker IFS66 v equipment (Bruker, US). Data were collected in absorbance mode using a frequency range of 4000–400 cm^{-1}, and resolution of 4 cm^{-1} (mid infrared region) with 250 co-added scans. Degree of methylation was determined as the average of the ratio of the peak area at 1747 cm^{-1} (COO-R) over the sum of the peak areas of 1747 cm^{-1} (COO-R) and 1632 cm^{-1} (COO—) as described earlier (Singthong et al., 2004).

2.4.3. Gas chromatography −FID

2.4.3.1. Sample preparation.. To determine the monomeric composition, artichoke by-product powder was hydrolysed using sulfuric acid (H_2SO_4) following the methods of Seaman (Yapo & Koffi, 2008) and trifluoroacetic acid (TFA) (Garna et al., 2006). Briefly, in the first method 35 mg of sample were dispersed in 300 μL of 72% H_2SO_4 at room temperature for 3 h, diluted with 1 M H_2SO_4 and heated at 100 °C for 2.5 h. Supernatant was dried at 38–40 °C in a rotary evaporator for chromatographic analysis. For hydrolysis with TFA 30 mg of sample were hydrolysed with 1.5 mL of TFA 2 M at 110 °C under inert conditions for 4 h (Garna et al., 2006).

In order to determine the monomeric composition of extracted artichoke pectin, powder samples were hydrolysed with TFA following the method mentioned above. Similarly, standards of D-arabinose, L-rhamnose, D-fructose, D-galactose, D-mannose, D-glucose and GalA at 20 mg mL^{-1} were hydrolysed to correct the obtained results.

Besides, enzymatic hydrolysis of artichoke pectin, obtained in the optimal conditions, has been carried out to determine GalA content. Samples (2% w/v) were dissolved in 0.05 M sodium acetate buffer (pH 5.0) and hydrolysed with 90 U mL^{-1} of Viscozyme® L preparation (Garna et al., 2006). Then, samples were incubated in an orbital shaker at 50 °C and 750 rpm (0.94 *g*) for 24 h and immediately immersed in boiling water for 5 min to inactivate the enzyme.

2.4.3.2. Chromatographic analyses.. Monomeric composition of artichoke by-product powder and extracted pectins was determined by GC-FID as trimethyl silylated oximes (TMSO) formed following the method of Brobst and Lott (1966). First, samples containing between 2.5 at 5 mg of sugars were added 0.4 mL of internal standard (I.S.) solution (0.5 mg mL^{-1} phenyl-β-glucoside). Afterwards, the mixtures were dried at 38–40 °C in a rotary evaporator. Sugar oximes were formed by adding 250 μL of hydroxylamine chloride (2.5%) in pyridine and heating the mixture at 70 °C for 30 min and then silylated with hexamethyldisilazane (250 μL) and TFA (25 μL) and kept at 50 °C for 30 min. Reaction mixtures were centrifuged at 6708*g* for 2 min at room temperature. Supernatants were injected or stored at 4 °C prior to analysis.

GC-FID analysis of TMSO derivatives was performed on an Agilent Technologies 7890A gas chromatograph (Wilmington, DE, USA) using a fused silica capillary column DB-5HT, bonded, crosslinked phase (5% phenyl-methylpolysiloxane; 15 m × 0.32 mm i.d., 0.10 μm film thickness) (J&W Scientific, Folson, California, USA). The oven temperature was initially 150 °C increased at a rate of 1 °C min^{-1} to 165 °C, then increased at a rate of 10 °C min^{-1}–200 °C, then increased at a rate of 50 °C min^{-1} to 380 °C and held for 2 min. The injector and detector temperatures were set at 280 and 385 °C, respectively. Injections were carried out in split mode (1:5) using nitrogen at 1 mL min^{-1} as carrier gas. To study the response factors relative to the internal standard, solutions containing arabinose, rhamnose, galactose, mannose, glucose, GalA, raffinose and stachyose were prepared over the expected concentration range in samples. The identities of carbohydrates were confirmed by comparison with relative retention times of standard samples. Data acquisition and integration were performed using Agilent ChemStation Rev. B.03.01 software. All analyses were carried out in duplicate and data were expressed as mean ± standard deviation (SD).

2.4.4. High performance size-exclusion chromatography with evaporative light scattering detector (HPSEC-ELSD)

The molecular weight distribution of carbohydrates present in the reaction mixtures obtained by Celluclast activity over artichoke by-product was determined by HPSEC-ELSD. Samples (0.65 mg mL^{-1}) were dissolved in water, filtered using a 0.45 μm syringe filter (Symta, Spain) and analysed in an Agilent Technologies 1220 Infinity LC System (Böblingen, Germany) equipped with an evaporative light scattering detector (1260 Infinity ELSD). The nebulisation temperature was 75 °C. Synthetic air at 85 °C was used for the evaporation at a flow-rate of 1.2 mL min^{-1}. The separation of carbohydrates was carried out on a TSK-GEL G5000PW_{XL} column (300 mm × 7.8 mm, 10 μm particle size) and TSK-GEL G2500PW_{XL} column (300 mm × 7.8 mm, 6 μm particle size) in combination with a TSK-GEL PW_{XL} guard column (40 mm × 6 mm, 12 μm particle size) (Tosoh Bioscience, Montgomeryville, PA, USA) using 0.01 M NH_4Ac, as mobile phase and elution in isocratic mode at a flow rate of 0.5 mL min^{-1} for 50 min. The column was thermostatised at 25 °C and the injection volume was 50 μL (~32 μg of total carbohydrates). Data acquisition and processing were performed employing Agilent ChemStation software.

Molecular weight of carbohydrates was calculated by the external calibration method using solutions of commercial pullulan standards (M_w 0.342–788 kDa) (Fluka Analytical) in the range 10–2250 mg L^{-1}. All analyses were performed in duplicate, obtaining relative standard deviation (RSD) values below 10% in all cases.

2.5. Statistical analyses

Experimental design, RSM and statistical analysis were performed using the software STATGRAPHICS Centurion XVI.I. (Statistical Graphics Corporation, Rockville, MD, USA). Other data not obtained in the experimental design were submitted to an analysis of variance ($P < 0.05$).

Table 1
Characterisation of powder artichoke by-product and extracted pectin.

	Artichoke by-product powder Mean ± SD	Extracted artichoke pectin Mean ± SD
pH	4.9	6.0
Dry matter (DM) (%)	91.7 ± 2.5	90.5 ± 0.3
a_w	0.20 ± 0.01	0.26 ± 0.00
Protein (mg g^{-1} DM)	140.5 ± 15.8	20.7 ± 1.0
Fat (mg g^{-1} DM)	2.4 ± 0.2	–
Alcohol insoluble fraction (AIF) (mg g^{-1} DM)	928.2 ± 5.9	–
Alcohol soluble fraction (ASF) (mg g^{-1} DM)	71.8 ± 5.9	–
Water insoluble fraction (WIF) (mg g^{-1} DM)	926.5 ± 1.6	–
Water soluble fraction (WSF) (mg g^{-1} DM)	73.5 ± 1.6	–
K (mg g^{-1} DM)	16.9	10.6
Na (mg g^{-1} DM)	3.5	9.6
Ca (mg g^{-1} DM)	4.7	7.3
Mg (mg g^{-1} DM)	2.4	2.9
B (mg g^{-1} DM)	0.02	0.04
Degree of methylation (%)	9.1 ± 0.5	19.5 ± 0.0

3. Results and discussion

3.1. Characterisation of artichoke by-product

The values of pH, dry matter (DM), a_w, protein, fat, degree of methylation, minerals and different carbohydrate fractions (AIF, ASF, WIF and WSF) found in industrial artichoke by-product powder are listed in Table 1. Protein, fat and mineral contents were similar to reported by USDA (2017) for artichokes cooked, boiled and drained.

Water and alcohol insoluble fractions (WIF and AIF), represented 926 and 928 mg g^{-1} DM. These values were higher than that found by Femenia et al. (1998) (64% for external bracts), however this difference could be explained by the possible leaching occurred during artichoke by-product blanching, as was indicated by Gamboa-Santos, Montilla, Soria, and Villamiel (2012) that reported a loss of weight up to 31% of soluble solids in the blanching water of carrot samples subjected to treatment at 95 °C during 5 min. WSF and ASF represented 74 and 72 mg g^{-1} DM of the sample, respectively.

Carbohydrate composition of WSF and ASF was determined by GC-FID, and as it can be seen in Table 2 the most abundant low molecular weight carbohydrates were fructose, glucose, sucrose, kestose and nystose, most of them probably released by the spontaneous hydrolysis of inulin, occurred during extraction of WIF and AIF.

Besides, artichoke by-product powder was hydrolysed with acids to know monomeric composition; the results are depicted in Table 2. As observed hydrolysates using H_2SO_4 (Seaman hydrolysis) contained arabinose, rhamnose, galactose and galacturonic acid (GalA) as pectic monosaccharides; and fructose, mannose and glucose indicating the presence of other polysaccharides as fructans, mannans, cellulose and hemicellulose in by-products from artichoke. The same type of monosaccharides was found in TFA hydrolysates, but in higher amounts than in H_2SO_4 ones, except for fructose and glucose. GalA levels in TFA hydrolysates were about 34 mg g^{-1} DM, fifteen times higher than using H_2SO_4. Since GalA is the main constituent of pectin, the high content found in TFA hydrolysates may indicate that the artichoke by-product is a good source to obtain pectin. In addition, taking into account the results of both hydrolysis, acid hydrolysis with TFA was selected to determine the monomeric composition of pectin extracted in the different assays of the experimental design with the aim of optimise the conditions of enzymatic extraction of artichoke pectin.

Table 2
Carbohydrate composition found in water (WSF) and alcohol soluble fractions (ASF) from artichoke by-product powder and hydrolysates of artichoke by-product powder using H_2SO_4 (Seaman hydrolysis) and TFA (2 M).

Carbohydrates (mg g^{-1} DM)				
	Water soluble fraction (WSF)	Alcohol soluble fraction (ASF)	Hydrolysates	
			H_2SO_4	TFA
Arabinose	–	–	14.2 ± 4.4	72.0 ± 18.5
Rhamnose	–	–	1.1 ± 0.0	10.9 ± 2.2
Fructose	21.8 ± 0.0	18.5 ± 1.1	4.4 ± 4.4	3.3 ± 2.2
Galactose	1.1 ± 0.0	–	2.2 ± 1.1	17.4 ± 4.4
Mannose	6.5 ± 1.1	4.4 ± 1.1	1.1 ± 0.0	29.4 ± 6.5
Glucose	15.3 ± 1.1	14.2 ± 1.1	19.6 ± 6.5	15.3 ± 4.4
Galacturonic acid	–	–	2.2 ± 0.0	33.8 ± 7.6
Sucrose	6.5 ± 1.1	7.6 ± 0.0	–	–
Kestose	6.5 ± 0.0	6.5 ± 0.0	–	–
Nystose	8.7 ± 1.1	12.0 ± 1.1	–	–

3.2. Optimisation of enzymatic extraction of pectin from artichoke by-product

In order to select the best enzyme to extract pectin from artichoke by-product, different commercial preparations were tested (Table S1). According to the results obtained, Celluclast presented high values of cellulase activity although pectinase activity was not detected, then it is expected that extracted pectin will not be degraded. On the other hand, Viscozyme®L presented the highest pectinase activity. Therefore, this enzyme was selected for enzymatic hydrolysis of pectin to determine GalA content of most accurate form.

Once Celluclast preparation was selected for pectin extraction, the experimental design was developed to optimize the extraction conditions using independent variables: artichoke by-product powder concentration (X_1), enzyme dose (X_2) and extraction time (X_3). Table 3 shows the results obtained during the enzymatic extraction of pectin from artichoke by-product for the different experimental variables studied. The yields of extracted pectin (Y_1) in the 17 assays of experimental design were in range of 98.1–208.3 mg g^{-1} DM. The highest pectin yield was obtained in a reaction with 4.5% of artichoke by-product powder, 7.8 U g^{-1} of enzyme during 27.2 h. This value was close to the theoretical optimum 225.7 mg g^{-1} DM obtained, according to the RSM analysis, after 27.2 h using 7.9% of substrate and 10.7 U g^{-1} of enzyme.

Following with the optimisation of pectin extraction design using Celluclast, other dependent variables, as GalA content (Y_2), pectic neutral sugars (Y_3, arabinose, rhamnose and galactose) and other neutral sugars (Y_4, fructose, glucose and mannose) were determined (Table 3). Hydrolysis of extracted artichoke pectin using TFA was performed to obtain individual values of each monosaccharide (see Table S2). The main carbohydrate quantified was GalA (Y_2) being in the range of 30.5–58.1 mg g^{-1} DM reaching the maximum value at the same conditions of substract concentration (4.5%) and enzyme dose

Table 3
Values of independent variables (artichoke by-product powder concentration X_1, enzyme dose X_2 and extraction time X_3) and dependent variables (pectin yield Y_1, galacturonic acid content Y_2, pectic neutral sugars Y_3 and other neutral sugars Y_4), and predicted and observed desirability found in each assay of pectin extraction after Celluclast treatment of artichoke by-products. Data are expressed as mg g^{-1} DM artichoke by-product powder.

Run	X_1, artichoke by-product powder concentration (%)	X_2, Enzyme dose (U g^{-1})	X_3, Extraction time (h)	Y_1, Pectin yield (mg g^{-1} DM)	Y_2, GalA (mg g^{-1} DM)	Y_3, Pectic neutral* sugars (mg g^{-1} DM)	Y_4, Other neutral** sugars (mg g^{-1} DM)	Predicted desirability	Observed desirability
1	7.0 (+1)	13.3 (+1)	24.0 (+1)	190.8	49.8 ± 2.4	32.6 ± 2.7	15.8 ± 1.0	0.90	0.80
2	2.0 (−1)	13.3 (+1)	6.0 (−1)	163.6	47.8 ± 0.1	21.3 ± 0.3	11.5 ± 0.7	0.34	0.27
3	4.5 (0)	7.8 (0)	15.0 (0)	176.7	52.2 ± 0.9	28.9 ± 0.5	15.0 ± 0.4	0.68	0.62
4	4.5 (0)	0.3 (−2)	15.0 (0)	106.9	***30.5 ± 0.5***	13.1 ± 0.1	***7.3 ± 0.2***	0.10	0.00
5	4.5 (0)	7.8 (0)	15.0 (0)	179.9	***58.1 ± 5.1***	31.2 ± 1.7	15.5 ± 1.7	0.68	0.70
6	4.5 (0)	7.8 (0)	15.0 (0)	197.4	50.6 ± 1.2	29.8 ± 0.7	16.2 ± 1.3	0.68	0.65
7	2.0 (−1)	2.2 (−1)	24.0 (+1)	133.0	45.3 ± 0.1	22.5 ± 0.4	10.0 ± 0.5	0.00	0.00
8	4.5 (0)	15.3 (+2)	15.0 (0)	187.6	51.6 ± 2.1	30.5 ± 2.1	16.4 ± 3.4	0.55	0.65
9	4.5 (0)	7.8 (0)	27.2 (+2)	***208.3***	57.6 ± 1.9	***37.2 ± 0.7***	***21.8 ± 1.2***	0.81	0.86
10	7.0 (+1)	2.2 (−1)	6.0 (−1)	***98.1***	32.2 ± 3.1	***12.2 ± 0.7***	8.2 ± 0.5	0.00	0.00
11	2.0 (−1)	13.3 (+1)	24.0 (+1)	167.9	52.1 ± 3.3	36.1 ± 2.8	12.6 ± 1.6	0.40	0.41
12	7.9 (+2)	7.8 (0)	15.0 (0)	185.4	48.7 ± 1.5	32.8 ± 0.1	21.5 ± 1.1	0.73	0.79
13	2.0 (−1)	2.2 (−1)	6.0 (−1)	159.2	37.4 ± 0.4	20.4 ± 0.4	9.7 ± 1.2	0.11	0.18
14	7.0 (+1)	13.3 (+1)	6.0 (−1)	129.8	44.6 ± 0.7	23.4 ± 1.1	13.4 ± 0.5	0.50	0.50
15	7.0 (+1)	2.2 (−1)	24.0 (+1)	165.8	38.6 ± 1.3	24.6 ± 1.0	13.8 ± 0.9	0.48	0.50
16	1.1 (−2)	7.8 (0)	15.0 (0)	169.0	51.6 ± 3.4	33.5 ± 0.9	11.2 ± 0.8	0.27	0.00
17	4.5 (0)	7.8 (0)	2.8 (−2)	158.1	41.9 ± 3.4	23.3 ± 1.4	12.6 ± 1.0	0.43	0.43

Y_2, Y_3 and Y_4: values found in TFA hydrolysates of extracted artichoke pectin (See Table S2).
* Σ of arabinose, rhamnose and galactose.
** Σ of fructose, glucose and mannose.

(7.8 U g^{-1}) which was found the highest yield but at lower time of extraction (15 h). The content of neutral sugars characteristic of pectin (Y_3) was between 12.2–37.2 mg g^{-1} DM; for these compounds maximum content was achieved at the highest extraction time and low artichoke by-product powder concentrations (4.5%). During the obtainment of pectin using Celluclast, other polysaccharides as cellulose, hemicellulose, mannan or inulin were extracted being the monosaccharides released from these polysaccharides glucose, mannose and fructose; the amount of these compounds (Y_4) in the different assays was between 7.3 and 21.8 mg g^{-1} DM. Unlike the other parameters, the value of this variable must be minimised.

The optimisation of the experimental conditions for pectin extraction and the results obtained of the most representative dependent variables under study (yield, GalA, pectic neutral sugars and other neutral sugars contents) were related by means of RSM. The regression equations describing the changes in these studied variables are shown in Table 4. As it can be observed, the variables that presented higher regression values were content of GalA, pectic neutral sugars, and pectin yield (R^2 = 93.9, 92.8 and 88.6%; and R^2_{adj} = 86.0, 83.6 and 73.8%, respectively). Moreover, these variables are the most important in a pectin extraction process; the extraction was aimed at maximizing pectin yield and to get a product with the highest content of pectic sugars (GalA, arabinose, rhamnose and galactose) that reflect a high purity of pectin. So, these variables were selected to optimize the enzymatic extraction process using desirability function. The corresponding three-dimensional representation of the desirability function obtained is shown in Fig. 1. Maximum predicted desirability (0.96) corresponds to an artichoke by-product powder concentration of 6.5%, an enzyme dose of 10.1 U g^{-1} and an extraction time of 27.2 h

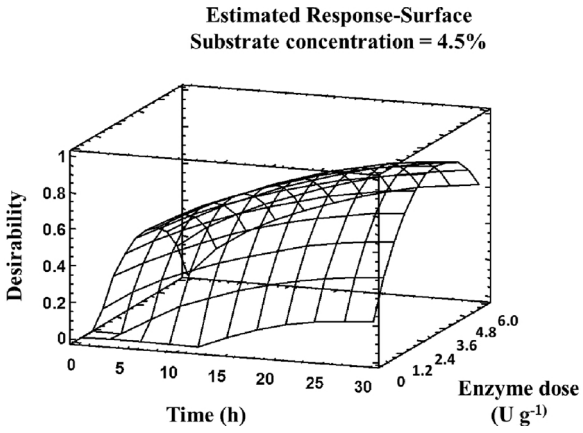


Fig. 1. Estimated response surface for the desirability function.

corresponding to a pectin yield of 175.6 mg g^{-1} DM. Among the tested conditions, the highest observed values of desirability function were 0.86, 0.80 and 0.79 (corresponding to assays 9, 1 and 12) and 0.66 ± 0.04 to the central points (assays 3, 5 and 6).

Taking into account that optimal extraction time (27.2 h) was the +α value given by the design and probably the maximum of pectin extraction may not have been reached, extraction time of pectin was increased at 36.0 and 48.0 h using the same optimised concentration of artichoke by-product powder (6.5%) and enzyme dose (10.1 U g^{-1}). According to the obtained results, pectin extracted after 48 h of hydrolysis contained more GalA and higher yields were obtained. Under the above mentioned conditions, the experimental yield was 22.1%

Table 4
Regression equations for the model fit of the different variables studied during the enzymatic extraction of pectin from artichoke by-product using Celluclast.

Variables	Fitted model equation*	R^2 (%)	R^2 adj (%)
Pectin yield (Y_1, %)	$Y_1 = 11.8809 + 1.22659X_2 + 0.0766667X_3X_1 - 0.069827X_2^2$	88.6	73.8
Galacturonic acid (Y_2, mg g^{-1})	$Y_2 = 0.593232 + 0.161405X_2 - 0.00790439X_2^2$	93.9	86.0
Neutral sugars from extracted pectin (Y_3, mg g^{-1})	$Y_3 = 0.647902 + 0.191896X_2 - 0.0110103X_2^2$	92.8	83.6
Other neutral sugars (glucose, mannose and fructose content, Y_4, mg g^{-1})	$Y_4 = 0.0661304 + 0.130052X_2 - 0.00695625X_2^2$	77.9	49.4

* (X_1) Artichoke by-product powder concentration (%), (X_2) enzyme dose (U g^{-1}), (X_3) extraction time (h).

(221.4 mg g^{-1} DM), which is well matched with the maximum predictive yield (225.7 mg g^{-1} DM).

Therefore, the use of enzymatic preparation Celluclast is a good method to extract pectin from artichoke by-products and it could be an alternative to traditional methods which use acids to extract pectin with high yield. Dominiak et al. (2014) extracted pectin from lime peel with different cellulases obtaining yield in the range of 18–26%, while using nitric acid the pectin yields were between 13 and 26%; similarly, Wikiera et al. (2015a) obtained higher yield with Celluclast (19%) than in sulfuric acid (15%) using apple pomace as raw material. Liew, Chin, Yusof, and Sowndhararajan (2016) obtained higher pectin yield from passion fruit peels with Celluclast (9.2%) than with citric acid (7.7%).

3.3. Characterisation of extracted artichoke pectin

Finally, a general characterisation of extracted artichoke pectin was performed. Thus, the physico-chemical parameters such as pH, DM, a_w, protein and degree of methylation were determined and the results are shown in Table 1. As can be observed the protein content of extracted pectin was low enough (20.7 mg g^{-1} DM), therefore it is not necessary to use any protease during enzymatic extraction to avoid presence of high level of protein in extracted pectin. With respect to mineral content in comparison with artichoke by-product, the most abundant cation was K, although its content was lower in the extracted pectin, besides, Na content was also greater probably due to buffer used during extraction. High Ca concentrations were found, probably due to high affinity of pectin to this divalent cation. Also, it is important to point out the high content of boron 0.04 mg g^{-1} DM, associated to the pectin (the content of artichoke by-product powder was 0.02 mg g^{-1} DM). In the bibliography has been reported that rhamnogalacturonan II (RG-II) suffer reversible dimerization with borate contributing to wall strength of the plant cell (Caffall & Mohnen 2009), similar role have calcium in apple and citrus pectin, while, in these products boron was not detected (Muñoz-Almagro, Montilla, Moreno, & Villamiel, 2017). The artichoke pectin obtained was classified as low-methoxyl pectin with a degree of methylation of 19.5%, similar to other pectins from sunflowers and lower than those of apple and citrus pectin previously characterised in our laboratory (72–76%).

Other parameter studied in extracted pectin was the Mw distribution analysed by HPSEC-ELSD. The pectin extracted at optimal conditions presented the profile with three fragments of 660, 105 and 4.8 kDa, with abundances of 28.5, 36.6 and 34.9% respectively (Fig. 2) This size distribution non monomodal is different to that of pectins extracted from citrus or apple (Muñoz-Almagro et al., 2017).

Following in artichoke pectin characterization, a determination of monomeric composition was performed in acid hydrolysates (TFA). Pectic monosaccharides such as arabinose (126.6 ± 5.8 mg g^{-1} DM of pectin), galactose (23.5 ± 0.4 mg g^{-1} DM), and rhamnose (18.8 ± 2.7 mg g^{-1} DM) were quantified. Besides, glucose (30.4 ± 4.4 mg g^{-1} DM), mannose (26.8 ± 0.8 mg g^{-1} DM) and fructose (4.6 ± 2.9 mg g^{-1} DM) were also present. Taking into account that GalA suffers a high degradation during acid hydrolysis, it was determined after enzymatic hydrolysis of pectin using Viscozyme® (Babbar et al., 2016). GalA content was 720.0 ± 44.6 mg g^{-1} DM (72.0% DM), which is higher than 65% being the limit to consider pectin as food additive (E-440) (Morris, Belshaw, Waldron, & Maxwell, 2013). The characteristic that could be highlighted of monomeric composition of extracted artichoke pectin is the high arabinose content. This is higher than found in pectin extracted from other sources like citrus and apple (Garna, Mabon, Wathelet, & Paquot, 2004; Gómez, Gullón, Yáñez, Parajó, & Alonso, 2013; Wikiera et al., 2015a) and similar to pectin from sugar beet pulp (Leijdekkers, Huang, Bakx, Gruppen, & Schols, 2015).

The amounts of rhamnose, arabinose and galactose with respect to GalA in the purified pectin indicated possible branching along the homogalacturonan (Yuliarti, Goh, Matia-Merino, Mawson, & Brennan, 2015). Considering the monomeric composition of pectin, the degree of branching (GalA/Rha), linearity pectin backbone [GalA/(Rha + Ara + Gal)] and extent of branching of RGI [(Ara + GalA)/Rha] were 19.2, 3.2 and 23.0, respectively. The degree and extent of branching of RGI were higher and the linearity pectin backbone was lower than those obtained for pectins from other sources as gold kiwifruit (Yuliarti et al., 2015) or lime peel (Dominiak et al., 2014). Another influential factor that can modify the structure of pectin is the extraction method, thus Fishman et al. (1991) obtained artichoke pectin with a low content of arabinose (3.4%), rhamnose (1.4%) and galactose (1.8%) using ammonium oxalate; being lower than orange pectin extracted with the same method, this fact may indicate a loss of branches. The maintenance of the structure, and more specifically the high arabinose content, could confer to this pectin anticarcinogenic property (Wikiera et al., 2015a).

As it has been indicated in the introduction, few papers have been published on the production and characterisation of artichoke pectin. Thus, Fishman et al. (1991) studied the monomeric composition of carbohydrate fraction extracted with ammonium oxalate from artichoke, found high amount of GalA and other typical monosaccharides of pectin. Femenia et al. (1998) reported GalA concentrations about 40% in the AIF of artichoke stems, shown to be a good source of pectic polysaccharide. Fissore et al. (2014) obtained soluble fibre extracts from artichoke by-products with a GalA content between 17 and 25% of total carbohydrates. However, no pectin separation was carried out in both cases.

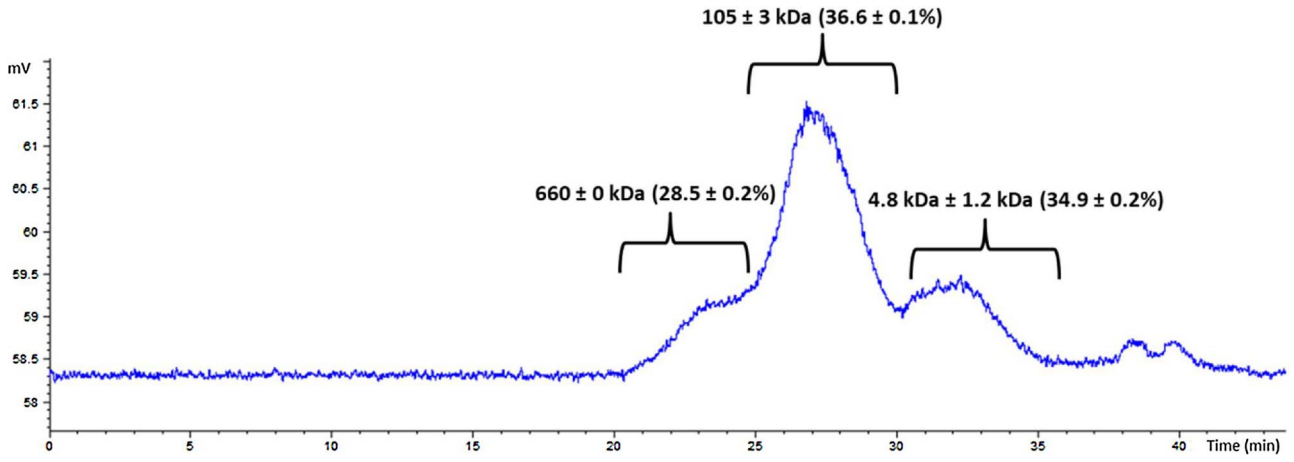


Fig. 2. HPLC-SEC-ELSD profile of a hydrolysate of artichoke by-product obtained by enzymatic hydrolysis using Celluclast® 1.5 L under optimal conditions (artichoke by-product powder concentration 6.5%; enzyme dose 10.1 U g^{-1}; extraction time 48 h).

4. Conclusion

Artichoke by-products (a mixture of stems, leaves and bracts) can be used as a good source of pectins. The extraction of pectin from these residues using commercial Celluclast preparation has been optimised allowing an efficient extraction giving rise to a yield of 221.4 mg g^{-1} DM. Enzymatic extraction is an environmentally-friendly process and allows to obtain high yields of pectin under optimal conditions. The pectin extracted from artichoke by-product powder was characterised finding a high GalA content (720.0 mg g^{-1}) and due to its degree of methylation (19.5%) was classified as low-methoxyled pectin. These characteristics would allow the use of this pectin as a food ingredient.

Acknowledgements

This work has been funded by MICINN of Spain, Project AGL2014-53445-R; ALIBIRD-CM S-2013/ABI-272, Comunidad Autónoma de Madrid and by the Spanish Danone Institute. Carlos Sabater thanks his FPU Predoc contract from Spanish MECD (FPU14/03619). Authors are also thankful to Riberebro (La Rioja, Spain) for kindly providing the artichoke by-products studied in this paper and Ramiro Martinez (Novozyme) for enzyme supply.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.carbpol.2018.02.055.

References

Association of Official Analytical Chemists (AOAC). (1990a). Method 920.151. In Association of Official Analytical Chemists (Eds.), Official methods of analysis of the AOAC (15th ed.). Arlington, VA: EEUU.

Association of Official Analytical Chemists (AOAC). (1990b). Method 920.152. In Association of Official Analytical Chemists (Eds.), Official methods of analysis of the AOAC (15th ed.). Arlington, VA: EEUU.

Association of Official Analytical Chemists (AOAC). (1990c). Method 920.39. In Association of Official Analytical Chemists (Eds.), Official methods of analysis of the AOAC (15th ed.). Arlington, VA: EEUU.

Babbar, N., Roy, S. V., Wijnants, M., Dejonghe, W., Caligiani, A., Sforza, S., et al. (2016). Effect of extraction conditions on the saccharide (neutral and acidic) composition of the crude pectic extract from various agro-industrial residues. *Journal of Agricultural and Food Chemistry, 64*(1), 268–276.

Brobst, K. M., & Lott, C. E. (1966). Gas-chromatographic investigation of hydroxyethyl amylose hydrolyzates. *Analytical Chemistry, 38*(12), 1767–1770.

Caffall, K. H., & Mohnen, D. (2009). The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydrate Research, 344*(14), 1879–1900.

Canteri, M. H. G., Nogueira, A., de Oliveira Petkowicz, C. L., Wosiacki, G. (2012). Characterization of Apple Pectin—A Chromatographic Approach. In *Chromatography-The Most Versatile Method of Chemical Analysis*. Chapter 14. InTech.

Dominiak, M., Søndergaard, K. M., Wichmann, J., Vidal-Melgosa, S., Willats, W. G., Meyer, A. S., et al. (2014). Application of enzymes for efficient extraction, modification, and development of functional properties of lime pectin. *Food Hydrocolloids, 40*, 273–282.

Femenia, A., Robertson, J. A., Waldron, K. W., & Selvendran, R. R. (1998). Cauliflower (*Brassica oleracea* L.), globe artichoke (*Cynara scolymus*) and chicory witloof (*Cichorium intybus*) processing by-products as sources of dietary fibre. *Journal of the Science of Food and Agriculture, 77*(4), 511–518.

Fishman, M. L., El-Atawy, Y. S., Sondey, S. M., Gillespie, D. T., & Hicks, K. B. (1991). Component and global average radii of gyration of pectins from various sources. *Carbohydrate Polymers, 15*(1), 89–104.

Fissore, E. N., Santo Domingo, C., Pujol, C. A., Damonte, E. B., Rojas, A. M., & Gerschenson, L. N. (2014). Upgrading of residues of bracts, stems and hearts of *Cynara cardunculus* L. var. scolymus to functional fractions enriched in soluble fiber. *Food & Function, 5*(3), 463–470.

Gamboa-Santos, J., Montilla, A., Soria, A. C., & Villamiel, M. (2012). Effects of conventional and ultrasound blanching on enzyme inactivation and carbohydrate content of carrots. *European Food Research and Technology, 234*(6), 1071–1079.

Garna, H., Mabon, N., Wathelet, B., & Paquot, M. (2004). New method for a two-step hydrolysis and chromatographic analysis of pectin neutral sugar chains. *Journal of Agricultural and Food Chemistry, 52*(15), 4652–4659.

Garna, H., Mabon, N., Nott, K., Wathelet, B., & Paquot, M. (2006). Kinetic of the hydrolysis of pectin galacturonic acid chains and quantification by ionic chromatography. *Food Chemistry, 96*(3), 477–484.

Gómez, B., Gullón, B., Yáñez, R., & Parajó, J. C. (2013). J.L. Alonso Pectic oligosacharides from lemon peel wastes: Production, purification, and chemical characterization. *Journal of Agricultural and Food Chemistry, 61*(42), 10043–10053.

Gullón, B., Gómez, B., Martínez-Sabajanes, M., Yáñez, R., Parajó, J. C., & Alonso, J. L. (2013). Pectic oligosaccharides: Manufacture and functional properties. *Trends in Food Science & Technology, 30*(2), 153–161.

López-Molina, D., Navarro-Martínez, M. D., Rojas-Melgarejo, F., Hiner, A. N., Chazarra, S., & Rodríguez-López, J. N. (2005). Molecular properties and prebiotic effect of inulin obtained from artichoke (*Cynara scolymus* L.). *Phytochemistry, 66*(12), 1476–1484.

Leijdekkers, A. G., Huang, J. H., Bakx, E. J., Gruppen, H., & Schols, H. A. (2015). Identification of novel isomeric pectic oligosaccharides using hydrophilic interaction chromatography coupled to traveling-wave ion mobility mass spectrometry. *Carbohydrate Research, 404*, 1–8.

Liew, S. Q., Chin, N. L., Yusof, Y. A., & Sowndhararajan, K. (2016). Comparison of acidic and enzymatic pectin extraction from passion fruit peels and its gel properties. *Journal of Food Process Engineering, 39*(5), 501–511.

Llorach, R., Espin, J. C., Tomás-Barberán, F. A., & Ferreres, F. (2002). Artichoke (*Cynara scolymus* L.) byproducts as a potential source of health-promoting antioxidant phenolics. *Journal of Agricultural and Food Chemistry, 50*(12), 3458–3464.

Machado, M. T., Eça, K. S., Vieira, G. S., Menegalli, F. C., Martínez, J., & Hubinger, M. D. (2015). Prebiotic oligosaccharides from artichoke industrial waste: Evaluation of different extraction methods. *Industrial Crops and Products, 76*, 141–148.

Maxwell, E. G., Belshaw, N. J., Waldron, K. W., & Morris, V. J. (2012). Pectin—An emerging new bioactive food polysaccharide. *Trends in Food Science & Technology, 24*(2), 64–73.

Miller, G. L. (1959). Modified DNS method for reducing sugars. *Analytical Chemistry, 31*(3), 426–428.

Min, B., Lim, J., Ko, S., Lee, K. G., Lee, S. H., & Lee, S. (2011). Environmentally friendly preparation of pectins from agricultural byproducts and their structural/rheological characterization. *Bioresource Technology, 102*(4), 3855–3860.

Mohnen, D. (2008). Pectin structure and biosynthesis. *Current Opinion in Plant Biology, 11*(3), 266–277.

Morris, V. J., Belshaw, N. J., Waldron, K. W., & Maxwell, E. G. (2013). The bioactivity of modified pectin fragments. *Bioactive Carbohydrates and Dietary Fibre, 1*(1), 21–37.

Muñoz-Almagro, N., Montilla, A., Moreno, F. J., & Villamiel, M. (2017). Modification of citrus and apple pectin by power ultrasound: Effects of acid and enzymatic treatment. *Ultrasonics Sonochemistry, 38*, 807–819.

Myers, R. H., Montgomery, D. C., Vining, G. G., Borror, C. M., & Kowalski, S. M. (2004). Response surface methodology: A retrospective and literature survey. *Journal of Quality Technology, 36*(1), 53–55.

Ruiz-Aceituno, L., García-Sarrió, M. J., Alonso-Rodriguez, B., Ramos, L., & Sanz, M. L. (2016). Extraction of bioactive carbohydrates from artichoke (*Cynara scolymus* L.) external bracts using microwave assisted extraction and pressurized liquid extraction. *Food Chemistry, 196*, 1156–1162.

Sałata, A., & Gruszecki, R. (2010). The quantitative analysis of poliphenolic compounds in different parts of the artichoke (*Cynara scolymus* L.) depending of growth stage of plants. *Acta Scientiarum Polonorum Hortorum Cultus, 9*(3), 175–181.

Santos, J. D. G., Espeleta, A. F., Branco, A., & de Assis, S. A. (2013). Aqueous extraction of pectin from sisal waste. *Carbohydrate Polymers, 92*(2), 1997–2001.

Schütz, K., Muks, E., Carle, R., & Schieber, A. (2006). Separation and quantification of inulin in selected artichoke (*Cynara scolymus* L.) cultivars and dandelion (Taraxacum officinale WEB. ex WIGG.) roots by high-performance anion exchange chromatography with pulsed amperometric detection. *Biomedical Chromatography, 20*(12), 1295–1303.

Singthong, J., Cui, S., Ningsanond, S., & Goff, H. (2004). Structural characterization, degree of sterification and some gelling properties of Krueo Ma Noy (*Cissampelos pareira*) pectin. *Carbohydrate Polymers, 58*, 391–400.

Terkmane, N., Krea, M., & Moulai-Mostefa, N. (2016). Optimisation of inulin extraction from globe artichoke (*Cynara cardunculus* L. subsp. *scolymus* (L.) Hegi.) by electromagnetic induction heating process. *International Journal of Food Science & Technology, 51*(9), 1997–2008.

USDA. (2017). United States Department of Agriculture. Agricultural Research Service USDA Food Composition Databases. https://ndb.nal.usda.gov/ndb/search/list.

Wikiera, A., Mika, M., Starzyńska-Janiszewska, A., & Stodolak, B. (2015a). Development of complete hydrolysis of pectins from apple pomace. *Food Chemistry, 172*, 675–680.

Wikiera, A., Mika, M., Starzyńska-Janiszewska, A., & Stodolak, B. (2015b). Application of Celluclast 1.5 L in apple pectin extraction. *Carbohydrate Polymers, 134*, 251–257.

Yapo, B. M., & Koffi, K. L. (2008). Dietary fiber components in yellow passion fruit rind: A potential fiber source. *Journal of Agricultural and Food Chemistry, 56*(14), 5880–5883.

Yapo, B. M. (2009). Biochemical characteristics and gelling capacity of pectin from yellow passion fruit rind as affected by acid extractant nature. *Journal of Agricultural and Food Chemistry, 57*(4), 1572–1578.

Yuliarti, O., Goh, K., Matia-Merino, L., Mawson, J., & Brennan, C. (2015). Extraction and characterisation of pomace pectin from gold kiwifruit (*Actinidia chinensis*). *Food Chemistry, 187*, 290–296.

Food Hydrocolloids 98 (2020) 105238

Contents lists available at ScienceDirect

Food Hydrocolloids

journal homepage: www.elsevier.com/locate/foodhyd

Ultrasound-assisted extraction of pectin from artichoke by-products. An artificial neural network approach to pectin characterisation

Carlos Sabater[a], Víctor Sabater[b], Agustín Olano[a], Antonia Montilla[a,*], Nieves Corzo[a]

[a] Instituto de Investigación en Ciencias de la Alimentación CIAL, (CSIC-UAM) CEI (UAM + CSIC), C/ Nicolás Cabrera, 9, E-28049, Madrid, Spain
[b] Sección de Ingeniería Química, Universidad Autónoma de Madrid, 28049, Madrid, Spain

ARTICLE INFO

Keywords:
Artichoke pectin extraction
Ultrasounds
Cellulase
Nitric acid
Sodium citrate
Artificial neural network

ABSTRACT

Artichoke (*Cynara scolymus* L.) by-products can be used as a good source of pectin. The aim of this work was to compare different pectin extraction methods: power ultrasound (US), enzymes, combination of US and enzymes (US + E), and acids (nitric and sodium citrate). After 6 h, pectin yield was higher when US was applied in combination with Celluclast®1.5 L (up to 13.9%). Structural characterisation showed that US-extracted pectins had lower weight-average molecular weight (M_w) values (146–155 kDa) than pectin extracted with US + E (160–267 kDa) and acid-extracted pectin (329–352 kDa). Monomeric composition reflected that pectin extracted with acids had the highest galacturonic acid (GalA) contents (82.2–90.2%) and the lowest degree of branching [Rha/GalA] (0.026–0.031). Structural characteristics of the different pectins were modelled using two artificial neural networks (ANN) considering composition parameters (Model I) and pectin FT-IR spectra (Model II). In addition, a third ANN was built to determine differences and similarities in the GC-MS spectra of monomeric composition (identified and unidentified monosaccharides) (Model III) showing characteristic patterns with high accuracy rates (above 95% on the test set). Structural differences depending on the extraction method of pectin have been established and these models could be applied to pectin from other sources.

1. Introduction

Pectins are complex polysaccharides widely used as functional food ingredient because of their technological and biological properties. Usually pectin is extracted from industrial by-products such as citrus peel, apple pomace and sugar beet pulp (Adetunji, Adekunle, Orsat, & Raghavan, 2017). However, nowadays there is an overall interest in finding alternative sources of pectin which could have different structures and enhanced physicochemical and biological properties. With this aim, non-traditional sources of pectins such as cranberry, onion, garlic, banana, mango, pumpkin, peach, rapeseed, papaya (Adetunji et al., 2017) or artichoke by-products (Sabater, Corzo, Olano, & Montilla, 2018a) have been studied.

The extraction process exerts a great influence on pectin structure. Chemical extraction employing strong acids such as hydrochloric, nitric and sulphuric acid and high temperature (70–100 °C) has been traditionally used in the industry (Adetunji et al., 2017; Maric et al., 2018; Yapo & Koffi, 2008). Nowadays, other food-grade extraction agents such as citrate are also studied (Adetunji et al., 2017; Marić et al., 2018; Muñoz-Almagro, Rico-Rodriguez, Wilde, Montilla, & Villamiel, 2018). Other extraction methods based on enzyme utilisation have been proposed as an environmentally-friendly alternative to acidic extraction with the advantage of the conservation of the original pectin structure (Marić et al., 2018). However, these methods can be highly time consuming, as observed by Sabater et al. (2018a) during pectin extraction from artichoke using Celluclast®1.5 L (Celluclast), a commercial cellulase preparation.

One promising and innovative extraction technique includes the assistance of power ultrasound (US) aimed at improving the yield, decreasing the energy consumption and shortening the processing time (Adetunji et al., 2017). The enhancement of extraction is mainly attributed to acoustic cavitation, generating microjets, shear forces and shockwaves at lower frequencies (20–100 kHz). Cavitation creates points with very high temperatures (which increase solubility and diffusivity) and pressures (which facilitate penetration and mass transfer) at the interface between a liquid medium and solid matrix subjected to ultrasonication; in addition, the microjetting and microstreaming effects cause the disintegration of solid materials and the disruption of cell walls. This phenomenon involves an increase in the contact between the solvent and the cell contents, enhancing mass transfer and increasing the extractive power (Chandrapala, Oliver, Kentish, & Ashokkumar, 2013; Soria, Villamiel, & Montilla, 2017). As a result, the

* Corresponding author. C/ Nicolás Cabrera 9, E-28049, Madrid, Spain.
E-mail address: a.montilla@csic.es (A. Montilla).

https://doi.org/10.1016/j.foodhyd.2019.105238
Received 11 January 2019; Received in revised form 28 June 2019; Accepted 19 July 2019
Available online 14 August 2019
0268-005X/

combination of US and enzymatic treatment can provide potential benefits. In general, it has been observed that, despite the reduction of the enzyme activity, the outcome of the combined process was positive (Nadar, Rao, Rathod, & 2018; Soria et al., 2017; Szabó & Csiszár, 2013).

Numerous authors studied procedures for ultrasound-assisted enzymatic extraction (UAEE) of biomolecules obtaining positive results. Thus, Wang, Sun, Liu, and Zhang (2014) extracted arabinoxylan, finding that the extraction yield noticeably increased with US power from 50 W (6%) to 200 W (12%), while power above 300 W resulted in a lower yield (6%). Similar results were obtained by Liu et al. (2014) during UAEE of polysaccharides from *Lycium barbarum,* obtaining under optimum conditions (20 min) a yield of 6.31%, which was almost similar to the yield of 6.81% obtained by enzyme-assisted extraction but much more effective due to the extraction time needed (91 min) at the same conditions. Similarly, Xu, Li, and Sun (2015) applied this combined technology to optimize the extraction of polysaccharides from blackcurrant. The influence of US waves on the activity and stability of enzymes has been shown to be specific for each enzyme and dependent on sonication parameters (Delgado-Povedano & de Castro, 2015; Rico-Rodríguez, Serrato, Montilla, & Villamiel, 2018).

On the other hand, several methodologies including multivariate data analysis may be used to study differences in pectin structure and composition. In general, data modelling allows valuable information to be discovered on chemical/structural properties and biological events (Käll, Canterbury, Weston, Noble, & MacCoss, 2007). For this purpose, machine learning algorithms may be employed. Supervised and semi-supervised classification methods support *a priori* known data structures to train patterns and rules to predict new data. Machine learning has been widely employed in the fields of metabolomics and food science (Lin et al., 2014; Sabater et al., 2018b; Uarrota et al., 2014). Some of their applications are product characterisation and component analysis (Fabris et al., 2010; Fernández Pierna et al., 2005), formulation of microparticles for polyphenol delivery (Belscak-Cvitanovic et al., 2015), detection of adulterations (Lim et al., 2017), ultrasonic extraction of phenolic compounds from wine lees (Tao, Wu, Zhang, & Sun, 2014) and study of starch gelatinization (Tao et al., 2018). In recent works, these algorithms have been used for a comprehensive characterisation of pectic oligosaccharide substructures (Sabater, Olano, Corzo, & Montilla, 2019; Sabater, Ferreira-Lazarte, Montilla, & Corzo, 2019a) and more studies dealing with the characterisation of larger pectin structures are needed.

These algorithms can also be applied for pattern recognition using mass spectrometry (MS) data (Jha, Josheski, Marina, & Hayashi, 2016), the preferred analytical technique to deepen knowledge on the structural properties of pectin and pectin derivatives (Ognyanov et al., 2016). To help spectra interpretation, other computational tools assisting structure elucidation have been developed, such as *in silico* fragmentation methods (Ruttkies, Schymanski, Wolf, Hollender, & Neumann, 2016; Sabater, Olano, Corzo, & Montilla, 2019b; Sabater, Ferreira-Lazarte, Montilla, & Corzo, 2019a). Among them, *Competitive fragmentation modelling* (CFM) generates MS fragmentation models for several techniques including GC-EI-MS, obtaining promising results (Allen, Greiner, & Wishart, 2015; Allen, Pon, Greiner, & Wishart, 2016).

Therefore, the aim of this work was to compare different pectin extraction methods, power ultrasound (US), enzymes, combination of US and enzymes (US + E), and acids (nitric and sodium citrate) from artichoke by-products. A structural characterisation using artificial neural networks (ANNs) to assess **I)** differences in pectin structural parameters (Model I), **II)** differences in pectin FT-IR spectra (Model II) and **III)** differences and similarities in the MS spectra of monomeric composition (identified and unidentified monosaccharides) of the various extracted pectins (Model III) has been carried out.

2. Materials and methods

2.1. Standards and samples

Analytical reference substances such as fructose, D-xylose, D-arabinose, L-rhamnose, D-galactose, D-mannose, D-glucose, galacturonic acid (GalA), and β-phenyl glucoside were purchased from Sigma Aldrich (Steinheim, Germany). Commercial multienzymatic preparations Celluclast®1.5 L (Celluclast, a cellulase from *Trichoderma reesei*) and Viscozyme®L (endo-1,3 (4)-β-glucanase from *Aspergillus aculeatus*), were generous gifts from Novozymes (Bagsvaerd, Denmark). Cellulase and pectinase activities were measured using the dinitrosalicylic acid (DNS) method. Industrial artichoke (*Cynara scolymus* L.) by-products were kindly supplied by Riberebro® (La Rioja, Spain).

2.2. Pectin extraction from artichoke by-products

Several extraction methods of pectin from artichoke by-products were used. All extraction experiences were made at least by duplicate.

2.2.1. Enzymatic extraction

Extractions were carried out in 50 mL polypropylene tubes with 15 mL of artichoke by-products suspension (6.5% w/v) prepared in sodium acetate buffer 50 mM pH 5.0 incubated with 10.1 U g^{-1} of Celluclast, with constant shaking (200 rpm), at 50 °C during 6 h. Pectin was purified according to Sabater et al. (2018a).

2.2.2. Ultrasound-assisted extraction

US treatments of artichoke by-product were carried out with Celluclast (10.1 U g^{-1}) (US + E) and without enzyme (US). The ultrasonic processor (450 Digital Sonifier, Branson Ultrasonics Corporation, Danbury, CT, USA) operated at a frequency of 20 kHz, with maximum power value of 400 W. A probe with a microtip horn of 3 mm diameter was immersed 2 cm with respect to the liquid surface into a 50 mL polypropylene tube (3 cm diameter, 11.5 cm height) with 15 mL of artichoke by-products suspension (3.25% w/v for US and 6.5% w/v for US + E treatments) suspended in sodium acetate buffer 50 mM pH 5.0. Pulsed US (2 s on/1 s off, US intensity value 81.7 W cm^{-2} corresponding to 30% amplitude) was the operating mode (Muñoz-Almagro, Montilla, Moreno, & Villamiel, 2017). The temperature (50 ± 0.1 °C) was registered with a temperature sensor (error ± 0.1 °C) and kept constant using an ice-water bath. The extraction processes were conducted with constant shaking (200 rpm) and several extraction times were studied (1, 2, 4 and 6 h). After US extraction, pectin was purified following the method of Sabater et al. (2018a).

2.2.3. Acid extraction

To extract pectin from artichoke by-products using acids (citric and nitric) the method of Muñoz-Almagro et al. (2018) for pectin extraction from sunflower by-products was followed.

Sodium citrate. 2 g of artichoke by-product were treated with sodium citrate 0.74% (5% w/v solids) at 72 °C for 3.23 h at constant shaking (150 rpm). The pH was adjusted to 3.25 with citric acid 2.5 M. The mixture was cooled in an ice-water bath and centrifuged at 3700*g* for 10 min. The precipitate was washed with two vol of acidified ethanol (0.2% HCl, v/v) and then centrifuged. The supernatants were collected, mixed and precipitated with two volumes of 96% ethanol (0.2% HCl v/v) and kept overnight at 4 °C. The solution was centrifuged, the supernatant discarded and the pellet washed with acidified ethanol (0.04% HCl, v/v; pellet/solvent ratio 1:2), centrifuged again and washed with ethanol 96% (pellet/solvent ratio 1:2), centrifuged and the drained precipitate was freeze-dried.

Nitric acid. 2 g of artichoke by-product were treated with nitric acid 0.24% (5% w/v solids, pH 2) at 90 °C for 3.95 h at constant shaking (150 rpm). The mixture was cooled in an ice-water bath and purified as

previously described for citrate extraction.

2.3. Analytical techniques

2.3.1. FT-IR spectra

In order to record pectin FT-IR spectra and to determine the degree of methyl-esterification (DM), freeze-dried samples of pectin extracted by the studied methods were analysed by FT-IR (Sabater et al., 2018a).

2.3.2. High performance size-exclusion chromatography with evaporative light scattering detector (HPSEC-ELSD)

Molecular weight (M_w) of extracted pectin was determined by HPSEC-ELSD following the method described by Sabater, Olano, Corzo, & Montilla, 2019b. The separation of carbohydrates was carried out on a TSK-GEL G5000PWXL column (300 mm × 7.8 mm, 10 μm particle size, molecular weight exclusion limit 1000 kDa) and TSK-GEL G2500PWXL column (300 mm × 7.8 mm, 6 μm particle size). V_0 (void volume) and V_T (total volume) of chromatographic system were 9.6 and 28.6 mL, respectively. Molecular weight of carbohydrates was calculated by the external calibration method using solutions of commercial pullulan standards (M_w 0.342–788 kDa) (Fluka Analytical). Weight average M_w of pectin was calculated as the weighted arithmetic mean of M_w of pectin fragments, considering their abundance (percentage): [Σ(M_w x abundance)/Σ(abundance)].

2.3.3. Gas chromatography coupled with mass spectrometry (GC-MS)

Monomeric composition of the different extracted pectins was studied by GC-MS. First, pectins (2% w/v in 0.05 M sodium acetate buffer; pH 5.0) were hydrolysed with 90 U mL^{-1} of Viscozyme®L preparation (Sabater et al., 2018a). Then, trimethyl silylated oximes (TMSO) of sugars released were analysed as previously described by Sabater, Olano, Corzo, & Montilla, 2019b. With monosaccharide data different ratios were calculated showing primary structural properties of pectin molecules, degree of branching (Rha/GalA), linearity pectin backbone [GalA/(Rha + Ara + Gal)] and extent of branching of RGI [(Ara + Gal)/Rha] (Wang et al., 2015; Yuliarti, Goh, Matia-Merino, Mawson, & Brennan, 2015).

2.4. Data analysis

ANOVA and Tukey's test for $p < 0.05$ were applied to all data generated.

Study of pectin composition parameters (Model I). Differences in extraction yield (%) and compositional parameters were used to classify pectin according to the extraction method through an artificial neural network (ANN) (Table 1, Model I). The ANN chosen was a multilayer perceptron (MLP) built with 1 hidden layer consisting in 7 neurons. The activation function was logistic. The architecture of MLP is shown in Fig. 1.

Study of pectin FT-IR spectra (Model II). A second MLP was computed to classify pectin according to the extraction method: using 709 wavenumbers in the range of 682–3600 cm^{-1} as inputs. These wavenumbers showed significantly different ($p < 0.05$) intensities among groups and corresponded to functional groups present in pectin structure. Before spectra classification, discrete wavelet transform (DWT) was applied to FT-IR data to increase MLP classification performance. A 14 level DWT was computed (indicating the depth of the decomposition) with a "la8" decomposition filter (Daubechies orthonormal compactly supported wavelet of length = 8). In the threshold step, DWT coefficients under percentile 15% were removed to denoise the signal (Sabater, Olano, Corzo, & Montilla, 2019b). The MLP used for this second study had 1 hidden layer consisting in 27 neurons.

Study of MS spectra of identified and unidentified monosaccharides present in pectin structures (Model III). A third study (Model III) comparing mass spectra of both identified and unidentified TMSO from sugars present in pectin and other polysaccharides structures co-

Table 1
Hydrolysates of artichoke pectin extracted with different methods classified considering different parmeters using artificial neural network (ANN) study (Model I). Rha: rhamnose, GalA: galacturonic acid, Ara: arabinose, Gal: galactose, Xyl: xylose.

Extraction method	Number of hydrolysates
Ultrasound (US) (1, 2, 4 and 6 h)	8 (2 per extraction)
Ultrasound and enzyme (US + E) (1, 2, 4 and 6 h)	8 (2 per extraction)
Enzyme (E) (6, 48[a] h)	4 (4 per extraction)
Sodium citrate (CIT) (3.23 h)	4 (2 per extraction)
Nitric acid (NIT) (3.95 h)	4 (2 per extraction)
Parameters studied in extracted artichoke pectins by ANN (Model 1)	
- Extraction yield (%)	
- Weight average molecular weight (M_w, kDa)	
- Galacturonic acid content (% total identified monosaccharides)	
- Pectic neutral sugar content (xylose, arabinose, rhamnose and galactose) (% total identified monosaccharides)	
- Non pectic monosaccharide content (fructose, mannose and glucose) (% total identified monosaccharides)	
- Degree of branching (Rha/GalA)	
- Linearity pectin backbone [GalA/(Rha + Ara + Gal)]	
- Extent of branching of rhamnogalacturonan I (RGI) [(Ara + Gal)/Rha]	
- Unidentified monosaccharide content (mg 100 mg^{-1} DM)	
- Degree of methyl-esterification (DM)	

[a] Sample obtained by Sabater et al. (2018a).

extracted during these treatments through a third MLP was carried out (Table 2). Only unidentified peaks (U) released during pectin hydrolysis which were not present in enzyme preparations were considered for the study. 75 fragments in the range of *m/z* 107–540 whose abundances were statistically different among groups ($p < 0.05$) were selected. These ions might correspond to monosaccharide fragments, assessed by the competitive fragmentation modelling source code (CFM-ID) developed by Allen et al. (2016). With this aim, complete feasible fragments for candidate molecules (Sabater, Olano, Corzo, & Montilla, 2019b) which could be present in pectin hydrolysates were calculated: identified monosaccharides, methyl-esterified GalA, acetylated GalA in O-2 or O-3, apiose, 3-deoxy-D-manno-octulosonic acid (KDO) and two phenolic compounds, coumaric and ferulic acid. Some RGIs (e.g., from Chenopodiaceae species such as spinach and sugar beet) are esterified with phenolics, including ferulic and/or coumaric acid, e.g., feruloylation at O-2, O-3, and O-5 of arabinose in α-L-arabinans and on O-6 of galactose in β-D-galactans (Atmodjo, Hao, & Mohnen, 2013). Ferulic and *p*-coumaric acids were considered because they are present in artichoke heads (Lattanzio & van Sumere, 1987).

Before MS spectra classification, a 14 level DWT was computed. The MLP used for this third study had 1 hidden layer consisting in 57 neurons. Fig. 2 shows its architecture.

Variables were scaled and centered before computing the analyses. All the models were trained with 70% of the data, 10-fold cross-validated and then tested with 30% of data from each class (corresponding to new samples). A variable importance analysis was carried out to determine the most influential parameters in the model. For this purpose, the sum of the product of raw input-hidden, hidden-output connection weights was calculated.

The three MLP developed were compared to two simpler models, Naïve Bayes classifier (NB) and Generalized Linear Model Elastic-net (GLMNET). NB is a simple classifier based on the Bayes Theorem that assumes that the presence of a particular feature (i.e. composition parameter, *m/z* ion) in a class (i.e. extraction method, GC-MS peak) is unrelated to the presence of any other feature. Considering these features, NB uses Naïve Bayesian equation to calculate the probability for each class. The class with the highest probability is the outcome of

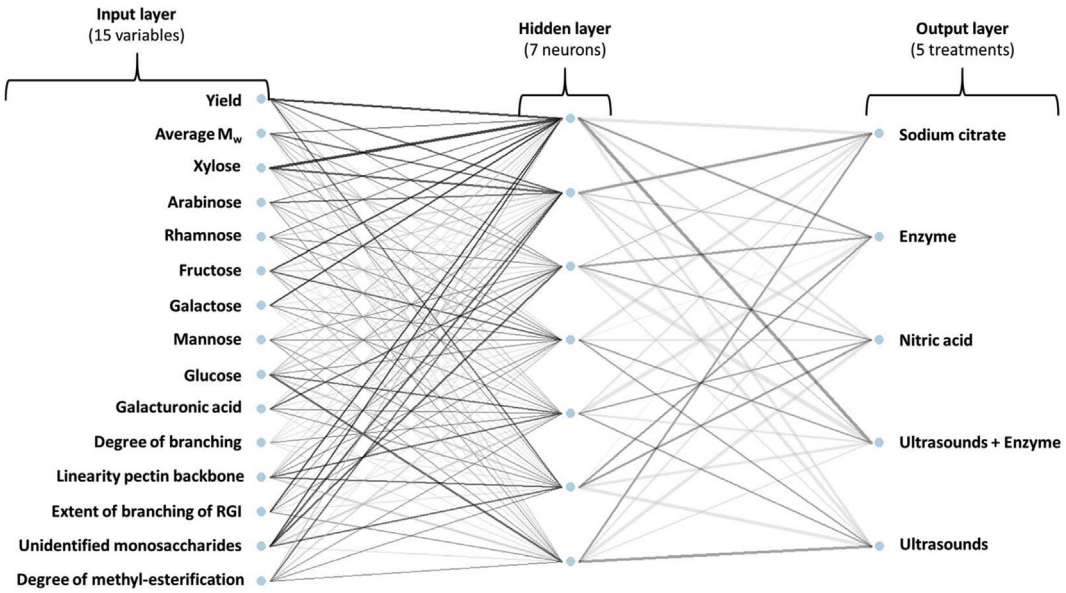


Fig. 1. Multilayer perceptron (MLP) architecture used for studying pectin structural parameters (Model I). Fifteen variables were selected as inputs. This model classified pectin samples according to their extraction method (acids, enzymatic, ultrasounds, and combination of ultrasounds and enzymatic treatments) as output. Weights are color-coded by sign (black +, grey -) and thickness is in proportion to magnitude. RGI: rhamnogalacturonan I. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Number of mass spectra of identified and unidentified monosaccharides (total n = 555) obtained from GC-EI-MS analysis of hydrolysates from artichoke pectins extracted following different methodologies included in the artificial neural network (ANN) study (Model III).

Compounds	Hydrolysates	Number of considered spectra	Number of samples	Total number of GC-MS spectra
Galacturonic acid	US, US + E, E, CIT, NIT	2	28 + 1 standard	58
Arabinose	US, US + E, E, CIT, NIT	1[a]	28 + 1 “	29
Fructose	US, US + E, E, CIT, NIT	2	28 + 1 “	58
Galactose	US, US + E, E, CIT, NIT	1	28 + 1 “	29
Rhamnose	US, US + E, E, CIT, NIT	2	28 + 1 “	58
Xylose	US + E, E	1	12 + 1 “	13
Mannose	US, US + E, E, CIT, NIT	1	28 + 1 “	29
Glucose	US, US + E, E, CIT, NIT	1	28 + 1 “	29
Unidentified monosaccharides (U) 1	US, US + E, E, CIT, NIT	1	28	28
2	US, US + E, E, CIT, NIT	1	28	28
3	US, US + E, E, CIT, NIT	1	28	28
4	US, US + E, E, CIT, NIT	1	28	28
5	US, US + E, E, CIT, NIT	1	28	28
6	US, US + E, E, CIT, NIT	1	28	28
7	US, US + E, E, CIT, NIT	1	28	28
8	US, US + E, E, CIT, NIT	1	28	28
9	NIT	1	4	4
10	US	1	8	8
11	US	1	8	8
12	US + E	1	8	8

US: ultrasound extraction, US + E: extraction combining ultrasounds and Celluclast, E: enzymatic extraction with Celluclast, CIT: extraction with sodium citrate, NIT: extraction with nitric acid.

[a] Only 1 isomer of TMSO derivatives.

prediction. In these models the bandwidth value (smoothing parameter) was set at 1. On the other hand, GLMNET involves a linear regression model that could be generalized (i.e. the response variable may follow different distributions than Normal distribution). Regularization methods are used to reduce possible overfitting of Generalized Linear Models and to reduce variance of the prediction error. Elastic-net regularization is based on two parameters, alpha and lambda. Alpha (comprised between 0 and 1) is used to optimize the model and it indicates the combination of two different regularization techniques, L1 and L2. All these methods try to penalize the Beta coefficients of the regression to get the important variables. Lambda is the regularization parameter or penalty coefficient, and allows adjusting the prediction error. Alpha values were set at 0.1 for all three GLMNET models while lambda was 0.028, 0.000 and 0.010 for Models I, II and III, respectively.

All statistical analyses were computed on R v3.5.0. DWT was computed using wavelets package (Aldrich, 2013). MLP models were built using the RSNNS package (Bergmeir & Benitez, 2012). NB models were computed with naivebayes package (Majka, 2019). GLMNET models were built using glmnet package (Friedman, Hastie, & Tibshirani, 2010).

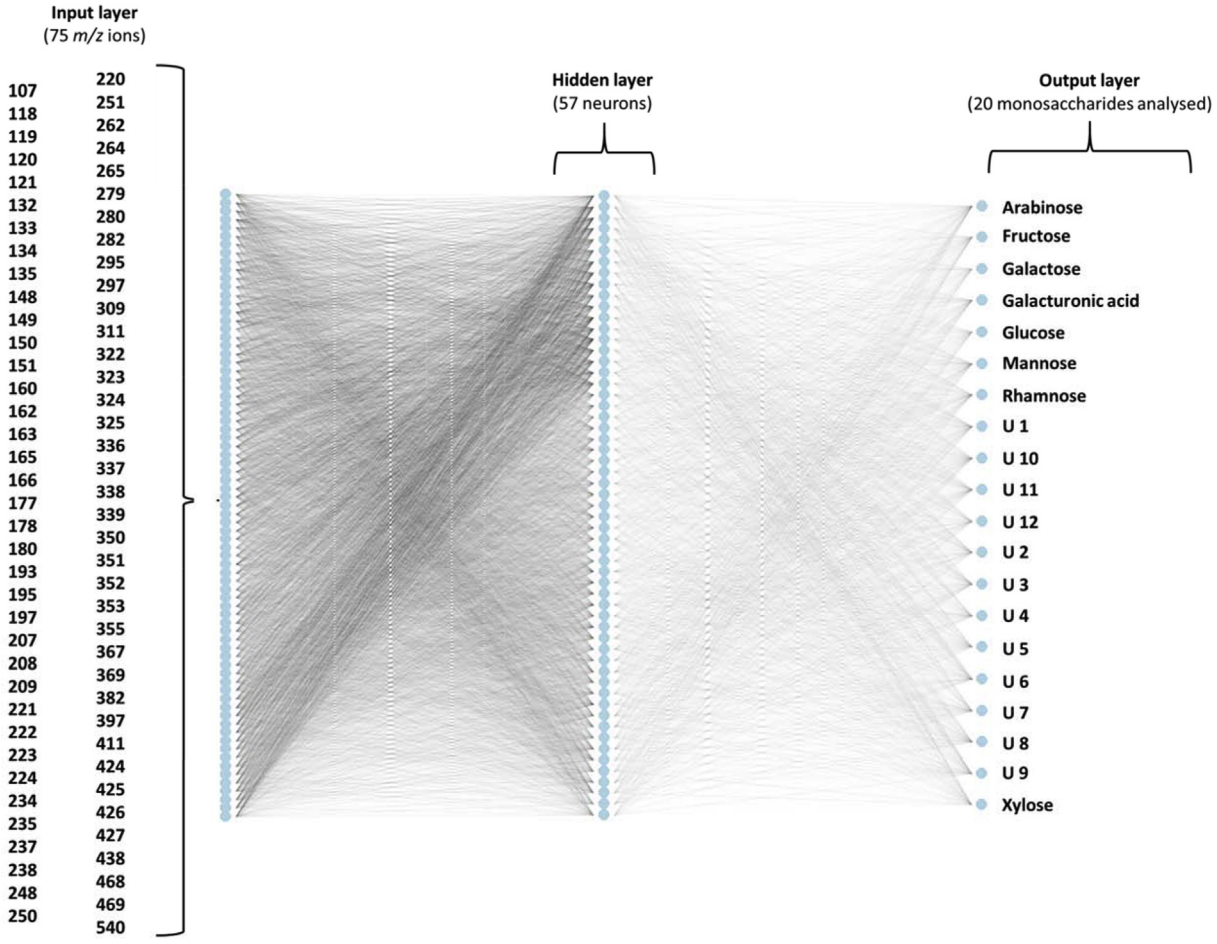


Fig. 2. Multilayer perceptron (MLP) architecture used for studying GC-MS spectra of identified and unidentified monosaccharides found in pectin hydrolysates (Model III). 75 ions in the range *m/z* 107–540 were selected as inputs. This model classified monosaccharide GC-MS spectra according to their chemical nature (8 identified and 12 unidentified monosaccharides, U) as output. Weights are color-coded by sign (black +, grey -) and thickness is in proportion to magnitude. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3. Results and discussion

The yield of pectin extracted from artichoke by-products following different methodologies is shown in Table 3. As can be seen lower pectin yield (7.1%) was obtained when US was applied to artichoke by-products compared to enzymatic extraction (11.3%) over 6 h. However, the highest extraction yield was achieved when US + E were used (13.9%). In contrast, conventional pectin extraction showed the lowest yields (4.8 and 5.8% for sodium citrate and nitric acid, respectively). The same trend was obtained by Li, He, Ly, & He (2014) who achieved the highest yield of water-soluble dietary fibre from apple pomace when they used enzymes and the lowest yield when acids were used, followed by US. With respect to conventional acid extraction, nitric acid has also been more efficient to extract pectin from sunflower heads (Muñoz-Almagro et al., 2018).

Numerous studies reported that US application enhances yields of polysaccharide extraction in general (Delgado-Povedano & de Castro, 2015; Ebringerová & Hromádková, 2010) and pectin extraction, in particular, i.e. from apple pomace (Panchev, Kirtchev, & Kratchanov, 1994), passion fruit peel (Dos Santos, Azoubel, & Gouveia, 2017), grapefruit peels (Wang et al., 2017), and sisal waste (Yang, Wang, Hu, Xiao, & Wu, 2018).

3.1. Characterisation of pectin extracted from artichoke by-products using different methods

Considering that enzymatic extraction of artichoke pectin using Celluclast preparation requires long times (48 h) (Sabater et al., 2018a), US extractions, with and without enzyme, were studied as an alternative method to reduce the extraction time. In addition, a comparison with conventional extraction methodologies using enzyme, nitric acid and sodium citrate has been carried out. Characterisation of extracted pectins, including yield, weight average M_w, monomeric composition and related parameters, and DM, is shown in Table 3.

Monomeric composition. Identified (n = 8) and unidentified (U, n = 12) monomeric compounds were found in GC-MS of hydrolysates of artichoke pectin extracted with the different methods assayed (Supplementary Material Fig. S1); quantitative data are compiled in Table 3. Pectin obtained with acids showed the highest GalA contents (90.2%), especially pectin extracted with nitric acid, similar to the results obtained by Muñoz-Almagro et al. (2018) to extract pectin from sunflower by-products (85.4–88.9%). Consequently, these pectins showed lower neutral sugar contents. On the contrary, low GalA contents were found in pectin extracted with US (61.6–68.3%) and especially in US + E treatments (59.3–66.4%). In both cases, GalA content decreased with US extraction time and after 6 h of the process, it was lower than 65%, the limit to consider pectin as a commercial food additive (E-440) (Morris, Belshaw, Waldron, & Maxwell, 2013). Regarding pectic neutral sugars, rhamnose ranged from 1.8 to 3.3% while

Table 3
Yield (mg of extracted pectin 100 mg^{-1} dry matter of artichoke by-product powder; %) and characterisation of artichoke pectin extracted using different methodologies: ultrasounds (US), combination of ultrasounds and Celluclast (US + E), Celluclast (E), sodium citrate and nitric acid.

Treatments (h)	Yield (%)	Average M_w (kDa)	GalA	Monomeric composition (% total identified monosaccharides)							Degree of branching Rha/GalA	Linearity pectin backbone GalA/(Rha + Ara + Gal)	Extent of branching (Ara + Gal)/Rha	Unidentified Mn (mg 100 mg^{-1} Dry Matter)	DM (%)
				Xyl	Ara	Rha	Fru	Gal	Man	Glc					
Sodium citrate 3.23	4.8^{e} (1.3)*	352^{a} (8)	82.19^{b} (2.54)	–	8.57^{d} (0.96)	1.75^{c} (0.54)	0.29^{g} (0.02)	2.75^{f} (0.16)	0.58^{c,d} (0.26)	5.03^{e} (0.29)	0.026^{d} (0.003)	5.85^{b} (0.44)	5.71^{e} (0.20)	4.37^{b,c,d} (0.53)	35.4^{a} (1.1)
Nitric acid 3.95	5.8^{d,e} (0.4)	329^{a,b} (14)	90.15^{a} (1.12)	–	0.95^{e} (0.22)	2.91^{a,b} (0.32)	0.47^{f} (0.03)	2.98^{e,f} (0.36)	0.44^{d} (0.10)	2.71^{f} (0.16)	0.031^{c,d} (0.003)	13.17^{a} (0.08)	1.48^{f} (0.23)	3.24^{d,e} (0.02)	34.5^{a} (1.8)
US 1	6.9^{c,d,e} (2.0)	155^{d,e} (25)	68.27^{c,d} (0.80)	–	14.03^{c} (0.99)	2.07^{b,c} (0.07)	0.50^{e,f} (0.02)	5.16^{b,c,d} (0.21)	1.60^{a,b} (0.23)	8.37^{c,d} (0.30)	0.030^{c,d} (0.001)	3.21^{c,d} (0.15)	9.28^{a} (0.69)	2.52^{e} (0.08)	20.9^{c} (0.6)
2	6.9^{c,d,e} (0.2)	152^{d,e} (14)	67.85^{c,d} (0.56)	–	14.01^{c} (0.82)	2.53^{a,b,c} (0.18)	0.57^{c,d,e,f} (0.03)	4.78^{c,d} (0.02)	0.84^{a,b,c,d} (0.02)	9.41^{b,c} (0.44)	0.037^{b,c} (0.003)	3.19^{c,d,e} (0.18)	7.43^{a,b,c,d} (0.21)	2.91^{d,e} (0.10)	22.0^{c} (1.9)
4	7.7^{c,d,e} (2.3)	153^{d,e} (0.2)	64.52^{d,e,f} (2.07)	–	15.08^{b,c} (1.26)	3.04^{a,b} (0.05)	0.63^{b,c,d} (0.04)	5.32^{a,b,c,d} (0.09)	0.94^{a,b,c,d} (0.13)	10.46^{a,b} (0.59)	0.047^{a,b} (0.001)	2.76^{d,e,f} (0.24)	6.72^{d,e} (0.55)	3.32^{d,e} (0.24)	22.8^{c} (1.0)
6	7.1^{c,d,e} (2.3)	146^{e} (6)	61.58^{e,f} (0.85)	–	15.71^{a,b,c} (0.25)	3.26^{a} (0.36)	0.69^{a,b} (0.04)	6.23^{a,b,c} (0.96)	1.03^{a,b,c,d} (0.08)	11.51^{a} (0.74)	0.053^{a} (0.007)	2.45^{e,f} (0.19)	6.76^{c,d,e} (0.37)	3.34^{d,e} (0.09)	29.7^{b} (1.6)
US + E 1	10.9^{b,c} (0.3)	160^{d,e} (6)	66.42^{d,e} (0.52)	0.53^{b} (0.13)	15.82^{a,b,c} (1.33)	2.61^{a,b,c} (0.47)	0.61^{b,c,d,e} (0.02)	5.54^{a,b,c,d} (0.88)	1.02^{a,b,c,d} (0.02)	7.46^{d} (0.26)	0.039^{b,c} (0.007)	2.77^{d,e,f} (0.13)	8.29^{a,b,c,d} (1.32)	3.46^{c,d,e} (0.18)	22.2^{c} (0.0)
2	12.4^{b,c} (1.8)	188^{c,d,e} (34)	64.35^{d,e,f} (0.90)	1.47^{a} (0.61)	16.23^{a,b,c} (0.05)	2.90^{a,b} (0.34)	0.68^{a,b,c} (0.03)	4.54^{d} (0.37)	1.44^{a,b} (0.62)	8.39^{c,d} (0.40)	0.045^{a,b} (0.005)	2.72^{d,e,f} (0.05)	7.21^{b,c,d,e} (0.70)	3.46^{c,d,e} (0.29)	22.2^{c} (0.8)
4	10.3^{b,c,d} (1.2)	267^{b,c} (59)	62.10^{e,f} (0.90)	1.51^{a} (0.04)	17.66^{a,b} (0.37)	3.05^{a,b} (0.18)	0.76^{a} (0.04)	4.80^{c,d} (0.03)	0.80^{b,c,d} (0.15)	9.32^{b,c} (0.53)	0.049^{a,b} (0.002)	2.43^{f} (0.06)	7.37^{b,c,d} (0.58)	3.73^{c,d,e} (0.35)	21.1^{c} (0.6)
6	13.9^{b} (0.1)	224^{c,d,e} (37)	59.28^{f} (1.62)	1.58^{a} (0.11)	17.87^{a} (1.10)	2.82^{a,b} (0.16)	0.75^{a} (0.02)	6.87^{a} (0.04)	1.61^{a} (0.01)	9.23^{b,c} (0.26)	0.048^{a,b} (0.004)	2.15^{f} (0.15)	8.77^{a,b} (0.11)	5.90^{b} (0.92)	22.8^{c} (1.7)
Enzyme 6	11.3^{b,c} (1.8)	206^{c,d,e} (4)	72.33^{c} (0.84)	0.47^{b} (0.00)	13.75^{c} (0.49)	2.09^{b,c} (0.09)	0.69^{a,b} (0.02)	4.21^{d,e} (0.36)	1.16^{a,b,c} (0.02)	5.30^{e} (0.62)	0.029^{c,d} (0.002)	3.61^{c} (0.08)	8.61^{a,b,c} (0.30)	10.37^{a} (1.75)	20.2^{c} (0.9)
Enzyme 48**	20.3^{a} (0.0)	228^{c,d} (5)	68.68^{c,d} (1.23)	0.53^{b} (0.19)	15.92^{a,b,c} (0.03)	3.25^{a} (0.18)	0.55^{d,e,f} (0.07)	6.38^{a,b} (0.57)	1.35^{a,b} (0.09)	3.33^{f} (0.42)	0.047^{a,b} (0.003)	2.69^{d,e,f} (0.13)	6.86^{c,d,e} (0.20)	5.49^{b,c} (0.02)	19.5^{c} (0.0)

^{a,b,c,d,e,f,g} Statistical differences between groups; Average M_w: weight average molecular weight determined by HPSEC-ELSD, GalA: galacturonic acid, Xyl: xylose; Ara: arabinose, Rha: rhamnose, Fru: fructose, Gal: galactose, Man: mannose, Glc: glucose, Unidentified Mn: unidentified monosaccharides including peaks U1-12 (Fig. S1). *: standard deviation. ** Data from Sabater et al. (2018a).

galactose and arabinose contents were higher in pectins treated with US and US with enzyme for prolonged times (6.2–6.9% and 15.7–17.9%, respectively). Xylose was found only in pectin enzymatically extracted (0.47%) and pectin treated with US and enzyme (0.5–1.6%). This pentose may be part of pectin, constituting the xylogalacturonan domain (Mohnen, 2008), or of hemicellulose, as xylan (Peng, Peng, Xu, & Sun, 2012); it is possible that the xylanase activity of Celluclast favours the release of xylan and its presence in the extracted polysaccharide (Berlin, Maximenko, Gilkes, & Saddler, 2007). With respect to non-pectic monosaccharides, glucose, mannose and fructose were high when US was applied reaching up to 11.5, 1.61 and 0.75% respectively. These results highlight the fact that US extraction processes are less selective and other polysaccharides like cellulose, hemicellulose and fructans may be extracted, increasing their presence with extraction time. This lack of selectivity was partially offset by the US + E combination, however cellulase activity was not enough to hydrolyse all cellulose.

This could be due to a loss of cellulase activity. Szabó and Csiszár (2013) reported that US power greatly affect the cellulase activity of Celluclast, so a treatment with probe (40 kHz, 500 W, 65 min) produced 12 and 25% of enzyme activity loss when amplitude was 40 and 80% respectively. However, the outcome of the enzyme-US combination was always positive. The advantageous effects of sonication on the heterogeneous systems are attributed to the increase in the mass transfer, due to acoustic cavitation, generating microjets, a strong physical agitation at the interface between a liquid medium and solid matrix (Soria et al., 2017). Also with Celluclast, Nguyen and Le (2013) observed that changes in ultrasonic intensity and sonication time significantly affected enzyme activity, increasing cellulolytic activity when ultrasonic intensity was increased from 0 to 6 W/mL and treatment time of 80 s, using a horn at 20 kHz, however the times were very short.

Finally, in hydrolysates of pectin an important group of unidentified monosaccharides (n = 12) was also detected (Table 2; Fig. S1). They were quantified together and amounts ranging from 2.5 to 10.4 mg 100 mg^{-1} dry matter were found (Table 3). They have been considered for the structural study of extracted polysaccharides.

Linearity and branching. Table 3 shows some ratios between contents of GalA and neutral pectic sugars in the purified pectins indicating possible branching along the homogalacturonan (Yuliarti et al., 2015). Considering the monomeric composition of pectin, the degree of branching (Rha/GalA), linearity pectin backbone [GalA/(Rha + Ara + Gal)] and extent of branching of RGI [(Ara + Ga)/Rha] were in the ranges 0.026–0.053, 2.2–13.2 and 1.5–9.3, respectively. The degree of branching was higher in pectin extracted with US and US + E due to its higher neutral sugar content, and lower in Celluclast and acid-extracted pectin because of its high GalA values. This parameter increased with prolonged US time as higher amounts of neutral sugars are released, while the linearity pectin backbone evolves inversely. The biggest differences among processed samples were observed in the other two pectin backbone parameters; pectins extracted with acids present a higher linearity and lower extent of branching of RGI with respect to the others, highlighting the structural variation of the pectins obtained and the importance of the RGI domain. From these molecular parameters, perhaps the most important for pectins could be the degree of branching. In general, this parameter for artichoke pectin (0.026–0.053) was similar for commercial apple pectin (0.033–0.037) (Muñoz-Almagro et al., 2017) and lower than for grapefruit peel pectin, extracted with or without US (0.12–0.13) highlighting the importance of RGI in this case (Wang et al., 2016).

FT-IR spectra. Degree of methyl-esterification. Fig. 3 illustrates FT-IR spectra of artichoke pectin extracted with different treatments, acids, US + E and E. Several bands corresponding to specific functional groups were determined, showing different intensities according to the extraction method. First, a significant band showing a wave number of 3300 cm^{-1} was observed in all samples. This band corresponds to O-H bonds stretch tension. This signal was high in acid-extracted samples and low in US + E treatments. On the other hand, the characteristic signal of C-H bonds stretch tension was found at 2920 cm^{-1} and was high in acid treatments. A signal corresponding to carboxylic acid C = O bonds stretch was reported at 1700 cm^{-1}. As expected, the highest intensities were reported in nitric-extracted pectin while intensity values found in US treatments were slightly higher than those found in the combination of US and enzymes. Carboxylate groups were found in the 1600 cm^{-1} band, with no significant differences among treatments. Other important signals were found at 1400 and 1200 cm^{-1}, corresponding to C-O-H bend deformation and carboxylic acid C-O stretch tension, respectively. In addition, several hydroxyl and ether characteristic peaks were found between 1000 and 1300 cm^{-1}. Finally, a 600-700 cm^{-1} signal was present in every sample, corresponding to the vibrational tension frequency of the pyranoid ring. This signal was high in enzymatically extracted pectin.

According to FT-IR spectra from artichoke pectin obtained by different methods could be classified as low-methyl-esterified pectin or pectic polysaccharide with a degree of methyl-esterification ranging from 20.2 to 35.4%, similar to other pectins from sunflowers (38–41%) and lower than those of apple and citrus pectin previously characterised in our laboratory (72–76%) (Muñoz-Almagro et al., 2017, 2018). DM was significantly higher in those pectins extracted with acids.

Molecular weight distribution. HPSEC-ELSD profiles of pectins extracted using different methods are shown in Fig. 4 and their quantitative data in Fig. 5. In general, pectin had a multimodal distribution and their patterns were not modified with the time of treatment. Pectin extracted with sodium citrate and nitric acid presented a similar bimodal pattern, with fragments > 600 kDa and around 130 kDa. US + E pectins presented three main fragments of > 600, 146–170 and 9–11 kDa. These profiles were similar to those obtained with Celluclast after 6 h of extraction that presented three fragments of 587, 130 and 11 kDa, and those obtained after 48 h with three fragments of 660, 105 and 4.8 kDa (Sabater et al., 2018a). US-extracted pectins showed a monomodal pattern, with a main fragment of 150–160 kDa and a small 9–15 kDa fraction, only corresponding at 2–5% (Fig. 5b). In general, the most common M_w distribution of pectins (from citrus, apple, sunflower) for HPSEC-ELSD analysis is monomodal (Muñoz-Almagro et al., 2017, 2018), however some pectins from sugar beet pulp show two peaks with different M_w (Yapo & Koffi, 2008), or from mango nectar a multimodal pattern (Huang et al., 2018).

With regard to weight average M_w of pectin (Table 3), US-extracted pectins showed lower values (146–155 kDa) than enzymatically obtained pectin (206 kDa), US + E (160–267 kDa) and acid-extracted pectin (329–352 kDa) possibly indicating that when only US was applied, under the conditions used, the treatment was not powerful enough to remove all of the cell-wall high M_w polysaccharides. This data agrees with that obtained by Yang et al. (2018), who explained that the higher molar mass attained from combined enzymatic-ultrasonic extraction can be explained as follows: firstly ultrasound is a relatively mild and not destructive extraction method in contrast to the acid methods. Secondly, only US did not release the larger calcium-bound pectin molecules which would be hard to extract only with this method. Interestingly, pectin extracted with acids showed higher weight average M_w values, because fragments between 2 and 20 kDa were not present. Also Wang et al. (2016) reported that US extracted pectin from grapefruit peel (20 kHz, 800 W, tip 25 mm, 2 s on/2 s off, US power density 0.41 W/mL, 67 °C over 28 min), has a lower M_w than that extracted without US (80 °C, 1.5 h), using deionised water at pH 1.5 adjusted with HCl as the extraction solvent. They explained that the US produced severe degradation. However, in our US-treated sample the fraction of low M_w was only 2–5%.

Overall, these results indicate that the application of different extraction methods using acids, US, Celluclast or US + Celluclast gave rise to pectins that were structurally different with respect to those obtained by conventional methods and with the possibility to use them as a functional ingredient and thus widen their application area.

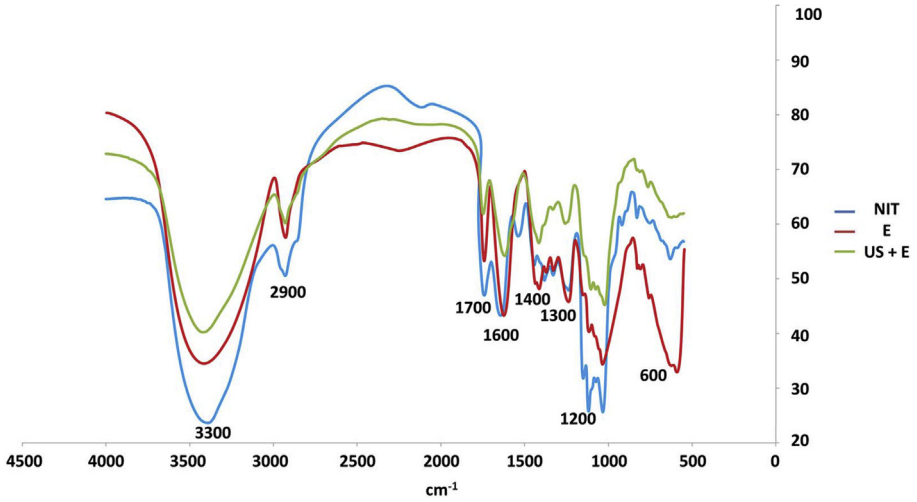


Fig. 3. FT-IR spectra of pectin extracted from artichoke by-products by different methods. NIT: nitric acid, E: enzymatic, US + E: combination of ultrasounds and enzymes.

3.2. ANN modelling of pectin compositional parameters

Structural differences and the composition of artichoke pectin were confirmed through data modelling (Model I, Table 1). Pectin samples were classified according to their extraction method (with nitric acid, sodium citrate, Celluclast, US and US + E) considering the parameters included in Table 3. With this aim, an ANN was applied (Fig. 1). This model was trained with 70% of data, 10-fold cross-validated and then tested on new samples (30% of data from each class i.e. extraction method). The training, cross-validation and test rates were 100, 91.67 and 100%, respectively. The overall kappa value was 0.86. This parameter is a more robust accuracy metric that considers the possibility of a correct classification by chance. The ANN model was compared to another two simpler models (Supplementary Fig. S2): NB (100, 79.19 and 100% training, cross-validation and test rates) and GLMNET (100, 85.00 and 100% training, cross-validation and test rates) that showed kappa values of 0.75 and 0.80. The model performance of ANN was similar to that of GLMNET. As these values indicate, this machine learning algorithm found a reproducible classification pattern for each kind of pectin that presented high prediction rates on the independent test set, showing no misclassification in this step. No misclassification occurred during the test phase so model sensitivity, specificity, true positive and negative rates and balanced accuracy were 100%.

The variable importance of ANN was calculated to determine the most influential parameters in this classification (Supplementary Material Table S1). The most influential variables for modelling were yield and xylose, glucose and unidentified monosaccharides. The yield showed great importance in classifying pectin extracted with Celluclast (importance coefficient, IC 6.53), also with positive values for US + E (IC 0.57). Xylose content was important probably due to the fact that it is only present in pectins extracted with enzyme and US + E (IC 1.46 and 7.34, respectively), while in the other pectins the coefficients were negative. High glucose content resulted to be relevant for classifying pectin extracted with US (IC 7.90), and may indicate that polysaccharides other than pectin are extracted during the process whereas low glucose contents were characteristic in nitric acid (IC -5.21) and Celluclast (IC -5.03) extracted pectins. Unidentified monosaccharide content was of great importance to differentiate artichoke pectin extracted with Celluclast (IC 5.53); this treatment produced the highest amount of these compounds. The other parameters have a minor individual relevance. In general terms, ANN modelling corroborates differences observed in experimental data and allows these products to be accurately classified according to the parameters studied. This model also highlights structural differences depending on the extraction

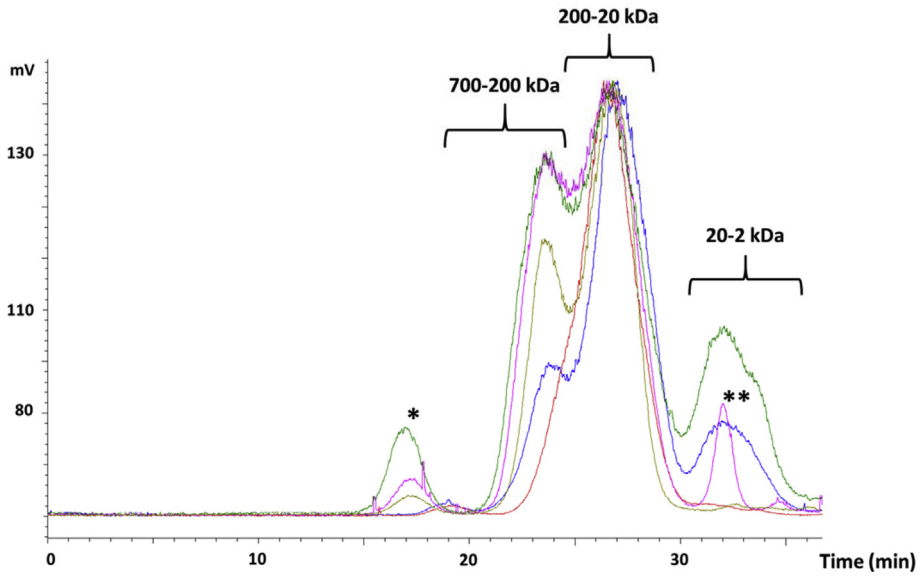


Fig. 4. HPSEC-ELSD profiles of artichoke pectin extracted from by-products using different methods: Celluclast for 6 h (blue), ultrasounds and Celluclast for 6 h (green), ultrasounds for 6 h (red), nitric acid (brown) and sodium citrate (pink). *Compounds out of M_w range **Sodium citrate peak. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

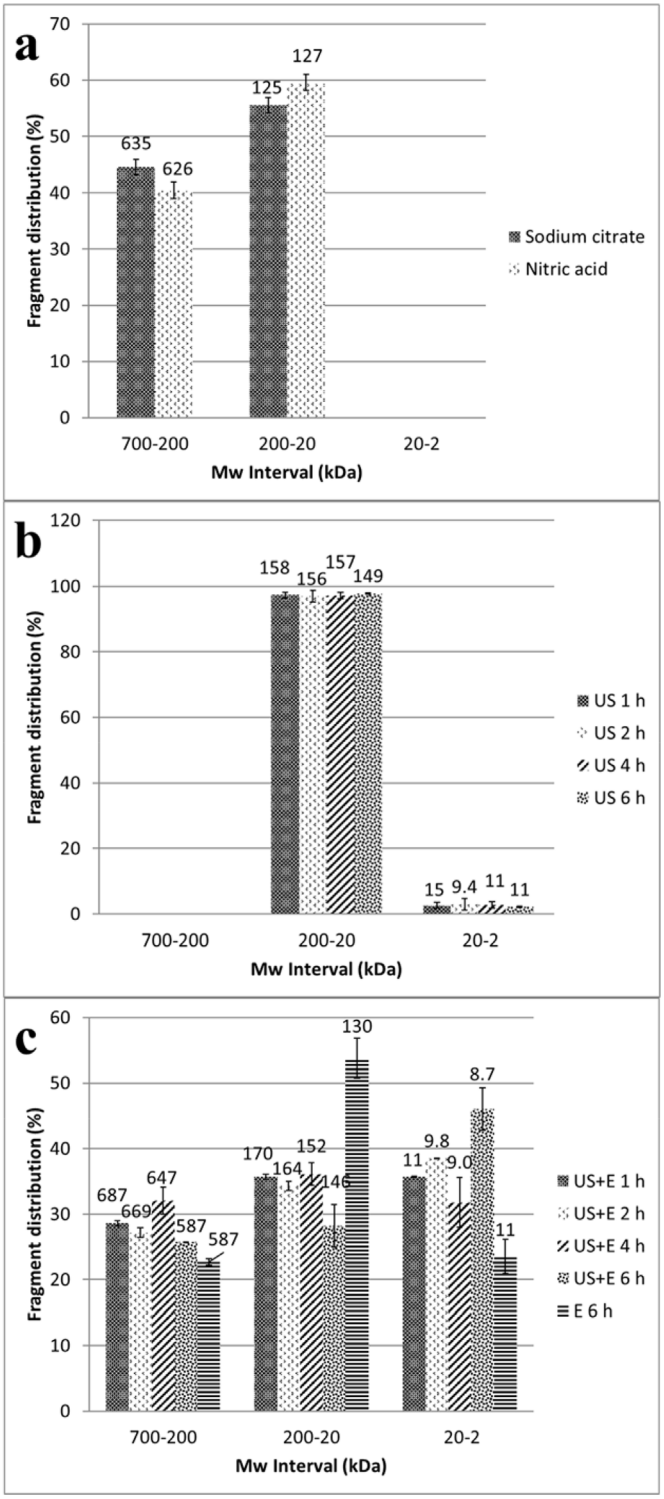


Fig. 5. Molecular weight distribution (HPSEC-ELSD) of artichoke pectin extracted using different methods: **a)** sodium citrate and nitric acid, **b)** ultrasounds (US) and **c)** Celluclast (E) and combination of ultrasounds and Celluclast (US + E).

method that may lead to different bioactivity. For example, high neutral sugar contents obtained in enzymatic treatments (including US + E), especially arabinose contents, could confer to this pectin an anticarcinogenic property (Wikiera, Mika, Starzyńska-Janiszewska, & Stodolak, 2015).

As a general overview, statistical differences among groups were calculated and they are shown in Fig. 6(a–e) considering extraction yield (%); weight average M_w (kDa); GalA content; pectic neutral sugar (xylose, arabinose, rhamnose and galactose) content; non-pectic monosaccharide (mannose and glucose) content. Also, statistical differences between parameters related to pectin chains were included (Fig. 6 f-h), considering degree of branching, linearity pectin backbone and extent of branching of RGI. As has been previously indicated, enzymatic extraction (with and without US) produced the highest yields. The lowest weight average M_w values were obtained in the US extraction. Pectic neutral sugar and other monosaccharide contents were higher in US-extracted pectin, with the latter showing a less selective extraction when US was applied. Finally, pectin extracted by acids showed high GalA content and linearity pectin backbone and a lower degree of branching and extent of branching of RGI, suggesting a less ramified structure with a different composition and different properties.

The statistical tests confirm the different patterns established during the supervised classification, which have a potential application on pectin from an unknown origin considering the high accuracy rates on new samples.

3.3. ANN modelling of pectin FT-IR spectra

Once it was demonstrated that pectin samples could be classified according to their composition parameters, these pectins were classified in a second study considering their FT-IR spectra (Model II). The training, cross-validation and test rates were 100, 83.15 and 100%, respectively. The overall kappa value was 0.82. The ANN model was also compared to another two models (Supplementary Fig. S3): NB (100, 79.94 and 100% training, cross-validation and test rates) and GLMNET (100, 80.59 and 100% training, cross-validation and test rates) that showed kappa values of 0.71 and 0.77. The model performance of ANN was similar to that of GLMNET. However, ANN performance was slightly higher than the rest of the models although these three algorithms made no misclassifications during the test phase indicating that a highly reproducible spectral pattern was found.

Similar to the previous model (Model I), model sensitivity, specificity, true positive and negative rates and balanced accuracy were 100%, indicating its high prediction rates. The most influential wavenumbers were then determined for each treatment (Supplementary Material Table S2). A total of 49 most relevant wavenumbers ranging from 1000 to 2828 cm^{-1} were selected to discuss statistically significant differences ($p < 0.05$) between FT-IR spectra of pectin extracted with different treatments:

Some wavenumbers (1000, 1019, 1024, 1042, 1044, 1055, 1063, 1081, 1083, 1103, 1104, 1131, 1142, 1145 and 1154 cm^{-1}) showed a high intensity in pectin extracted with acids and Celluclast and may correspond to the characteristic bend deformation of hydroxyl and ether groups. Other wavenumbers (1258, 1271, 1308, 1680, 1684, 1718, 1723, 1745 and 1762 cm^{-1}) were low in US treatments. As previously commented, wavenumbers 1258–1308 may correspond to hydroxyl and ether groups while wavenumbers 1680–1762 indicate carboxylic acid C=O bonds stretch. Some wavenumbers (2703, 2723, 2748, 2759 and 2772 cm^{-1}) were high in Celluclast treatments. Finally, other wavenumbers (1230, 1700, 1769, 1771, 1772, 1773, 1774, 1776, 1777, 1778, 1817, 1825, 1830, 1831, 1838, 2786, 2800, 2805, 2822 and 2828 cm^{-1}) were high only in citrate and Celluclast treatments; wavenumber 1230 cm^{-1} corresponds to hydroxyl and ether groups, 1700-1838 cm^{-1} to C=O groups, and 2786–2828 correspond to O-H bonds stretch.

This FT-IR spectra study confirms structural differences of pectin depending on the extraction method and gives complementary information to the first study based on composition parameters. On the other hand, it is important to note that, if desired, pectin could be classified according to the extraction method with this simple analysis, although deep structural information would not be obtained. To achieve this purpose other structural determinations, such as those discussed in the model I, and especially a study of spectra GC-EI-MS,

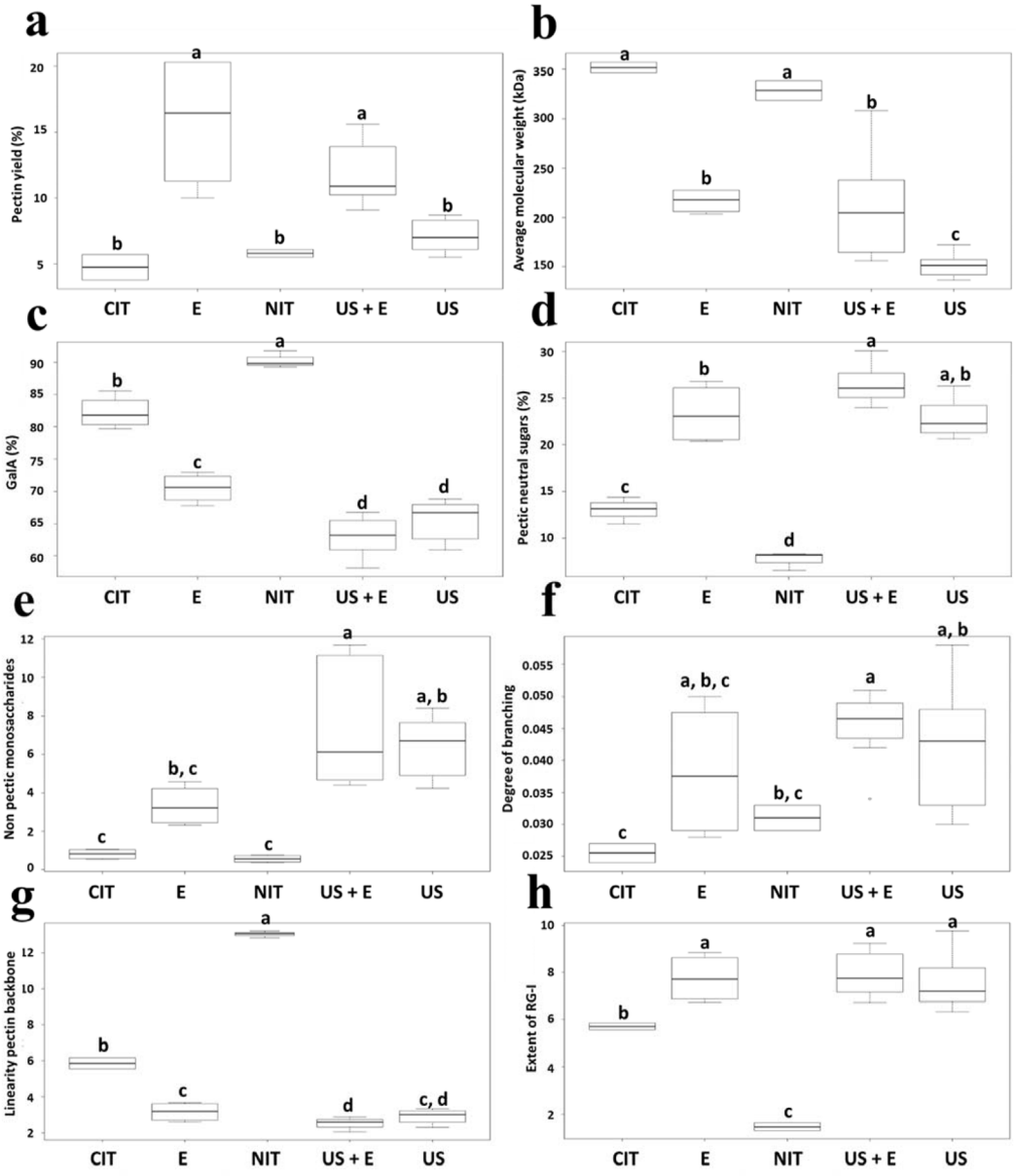


Fig. 6. Values of pectin extraction yield (%) (**a**), weight average molecular weight (kDa) (**b**), galacturonic acid (GalA) (c), pectic neutral sugar (xylose, arabinose, rhamnose and galactose) (**d**) and non pectic monosaccharides (mannose and glucose) (**e**) contents, degree of branching (Rha/GalA) (**f**), linearity pectin backbone [GalA/(Rha + Ara + Ga)] (**g**) and extent of branching of rhamnogalacturonan I [(Ara + Ga)/Rha] (**h**) values, found in pectin samples extracted by different methods, US: ultrasound, US + E: combining ultrasounds and Celluclast, E: Celluclast, CIT: sodium citrate, NIT: nitric acid. ^{a,b,c,d} Statistically significant differences ($p < 0.05$) between pectins.

made in the model III are necessary.

3.4. ANN study of GC-EI-MS spectra of identified and unidentified monosaccharide constituents of artichoke pectin

To gain deeper knowledge of the structures of extracted pectin, a GC-MS study of their monomeric composition, both of identified and unidentified sugars detected in the chromatographic profiles of hydrolysates (Supplementary Material Fig. S1) was carried out (Table 2, Model III). As can be observed, identified monosaccharides (xylose, arabinose, rhamnose, fructose, galactose, mannose, glucose and GalA) and 12 unidentified peaks were present in these structures. Unidentified peaks 1–8 (U1-U-8) were present in pectin extracted with all studied methods. U9 was found in nitric acid-extracted pectin while U10 and U11 were specific of pectin extracted with US and U12 was only present in pectin extracted with US + E (Table 2). Similarities and differences in the mass spectra of these identified and unidentified compounds were established using a third ANN (Model III). One standard for each identified monosaccharide was also included in the study, as reference. A total of 555 mass spectra were classified considering 75 to be relevant *m/z* ions (Fig. 2). To increase model performance, each MS spectrum was decomposed and reconstructed using the DWT. Then, ANN classified each spectrum according to each identified/unidentified peak. The training, cross-validation and test rates were 99.74, 97.19 and 96.97%, respectively. The overall kappa value was 0.95.

The most influential *m/z* ions in the classification were determined (Supplementary Material Table S3). In addition, model sensitivity, specificity, true positive and negative and balanced accuracy rates were calculated (Supplementary Material Table S4). Again, MLP found a reproducible classification pattern that showed high accuracy on the test set (above 95%). Interestingly, model sensitivity was 100% for the studied compounds with the exception of U7 (80.0%) galactose (81.8%)

and rhamnose (94.1%). Moreover, specificity rates were ≥98% in all cases. In general, true positive rates were 100% with the exception of rhamnose (94.1%), U11 (66.7%) and U6 (62.5%), probably due to structural similarities with other monomeric compounds that may interfere with the model predictions. However, true negative rates were ≥98% and the lowest balanced accuracy values found were 90.0 and 90.9% for U7 and galactose, indicating in all cases a high predictive power. Similar to the previous models, ANN was compared to other models (NB and GLMNET) that reached training, cross-validation and test rates of 97.69, 89.03, 95.76% (NB) and 98.46, 86.33, 93.33% (GLMNET). Kappa values were of 0.88 and 0.86 for NB and GLMNET, respectively. Although these simpler models allow GC-MS spectra to be classified accurately showing high sensitivity and specificity (Supplementary Table S5 and Table S6), some of the compounds studied could not be classified, especially unidentified compounds U9 and U12, probably due to the complexity of its spectra. Therefore, ANN outperformed the rest of the models when tested on new samples (Supplementary Fig. S4).

To deepen the study, possible chemical structures of studied fragments were suggested employing CFM-ID code (Allen et al., 2016). This code is able to calculate all possible fragments that can be formed from one molecule (the ones included in our *in silico* fragmentation library), including both immediate descendants and fragments formed from these descendants. Therefore, the chemical origin of relevant *m/z* ions has been determined (Supplementary Material Fig. S5). Proposed structures of the most interesting *m/z* ions (n = 27) are shown in Table 4. It should be noted that these structures contain nitrogen and trimethyl silyl groups because they correspond to TMSO fragments of monomeric compounds that could be used as marker ions. In addition, non-derivatized chemical substructures that are present in the original carbohydrate molecule are included in Table 4. Then, differences in the GC-MS spectra of identified and unidentified compounds were summarised, considering statistically significant differences ($p < 0.05$) in their abundances (Supplementary Material Fig. S6):

Relevant ions in neutral monosaccharides spectra. Some relevant ions (*m/z* 309 and 324) could be derived from the fragmentation of all identified monosaccharides, although their abundances were different. Specifically, the abundance (up to 5–10%) of ion *m/z* 309 was high in arabinose and xylose while the highest abundance (up to 5–10%) of fragment *m/z* 324 was observed in glucose and galactose. Other ions could be formed only from some monosaccharide including fragment *m/z* 323 (abundance around 1%) which could originate from glucose, galactose and rhamnose.

Relevant ions in unidentified peaks spectra. With regard to these peaks the abundances of other ions were important. The most relevant ions for some of these peaks and their possible origin are indicated below.

U2 spectra. The abundance of ion *m/z* 195 (up to 10%), which could be formed from methyl-esterified GalA, was high.

U7 spectra. The abundance of ion *m/z* 424 (around 5%), which could be formed from apiose, was high.

U8 spectra. Ions *m/z* 234, 248, 250, 262, 350 and 352 showed high abundances (between 1 and 10%). These fragments could be derived from KDO.

U9 spectra. The ions with high abundances were *m/z* 197, 235, 339 and 469 (between 1 and 5%). Fragment *m/z* 197 could be derived from ferulic acid, present in artichoke (Lattanzio & van Sumere, 1987) and in sugar beet pulp pectin (Yapo & Koffi, 2008). While ion *m/z* 469 could be also formed from the fragmentation of methyl-esterified/acetylated acidic sugars; ion *m/z* 235 could also be from rhamnose. Finally, fragment *m/z* 339 could also originate from all studied monosaccharides. U9 compound might be an acid molecule, maybe ferulic acid or an unknown acidic sugar. Probably, these two types of molecules coelute in the same U9 peak. This could explain the low accuracies and worse performances obtained for the classification of U9 spectra using simple models.

U11 spectra. The ion *m/z* 297 (abundance around 5%) could be formed from the fragmentation of methyl-esterified acidic sugars.

U12 spectra. Ions *m/z* 251, 325 and 355 were more abundant (abundances 1–5%). Ion *m/z* 251 could originate from rhamnose and *m/z* 325 and 355 from acidic sugars (acetylated or not).

U9 and U11 spectra. Important ions for both peaks were *m/z* 107 and 121 (abundances 0.1–0.5%). These two fragments could be derived from methyl-esterified acidic sugars, and ferulic/coumaric acid fragments.

U9 and U12 spectra. The abundance of ions *m/z* 209 (abundances between 0.1 and 0.5%) was higher in unidentified peaks 9 and 12 than in the rest of the peaks. This fragment could be derived from acidic sugar derivatives (methyl-esterified or not).

Relevant ions in several unidentified peaks spectra. Fragment *m/z* 133 (abundance 1–10%), also formed from rhamnose was high in U1, U2 and U11 peaks. Ion *m/z* 193 (abundance 1–5%) was high in U2, U5 and U12 peaks and could also originate from acidic sugars (methyl-esterified). Ion *m/z* 151 (abundance 1–3%), formed from acidic sugars, was high in U2, U7 and U9 peaks. Moreover, ion *m/z* 369 (high in U5 and U12; abundance 1–3%) could be formed from methyl-esterified acidic sugars. At last, ion *m/z* 411 (abundance 1–3%) was higher in U5 and U7 than in the rest of the peaks and could originate from apiose and most identified sugars.

These results suggest the presence of apiose (probably U7 peak), KDO (or similar molecules, probably U8 peak) and methyl-esterified (probably U2 and U11 peaks) or acetylated GalA or other acidic sugars (probably U1, U12 and other unidentified peaks) might be present in all pectin samples extracted by different treatments. It should be noted that the amounts of these unidentified monomers were higher in pectin extracted using enzymes (Table 3) and may be related with specific biological properties. This pectin also showed high amounts of known pectic sugars. Moreover, phenolic compounds like ferulic acid might be present in artichoke pectin, especially in pectin extracted with nitric acid (probably U9 peak), and might confer a potential antioxidant activity.

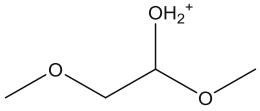
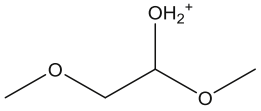
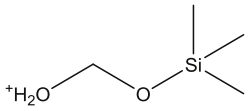
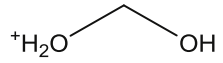
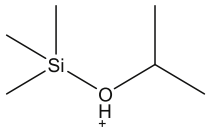
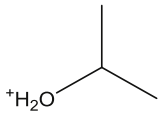
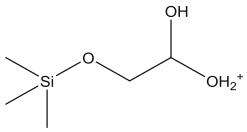
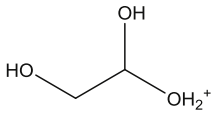
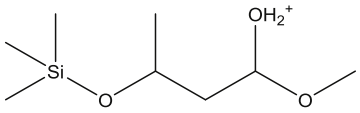
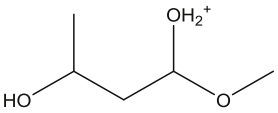
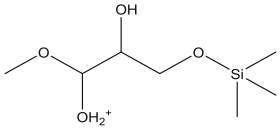
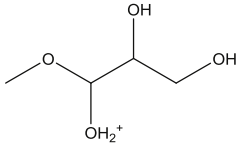
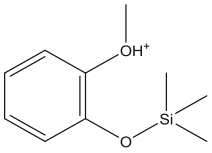
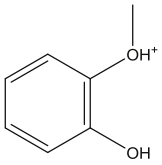
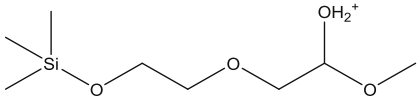
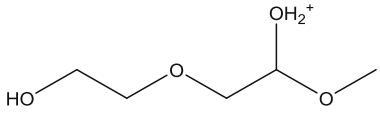
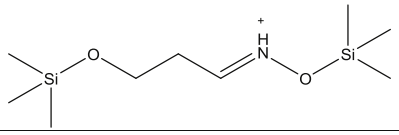
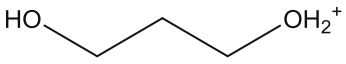
The ability to determine structural similarities and differences among pectins extracted by conventional and US-assisted methods as well as identified and unidentified compounds present in their structure could be of great importance to evaluate their functional activity. Given the great variability in the composition of the pectins according to the plant source and extraction methods, tools to classify these samples according to their structure may lead to a better understanding of their chemical properties. In this regard, further research is needed to establish structure-function relationships.

4. Conclusions

Different pectin extraction methods have been studied. In the extraction time studied, 6 h, US combined with Celluclast produced the highest yields compared to US alone; however, GalA content was lower than 65%, showing that US treatments produced the highest extraction of other polysaccharides. With respect to molecule structure, pectin obtained with US + E and with Celluclast showed similar linearity pectin backbone and extent of branching of RGI. Acid-extracted pectins presented a significantly higher GalA content. Taking into account these results a sequential combination of enzyme and US could be used to optimize the obtainment of pectin in short times, with high yields, low co-extraction of other polysaccharides, and without important molecular modification, except a lower GalA content.

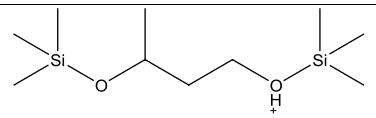
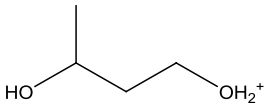
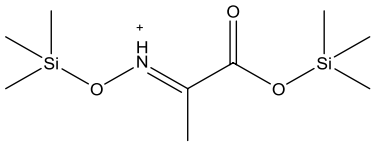
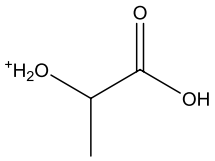
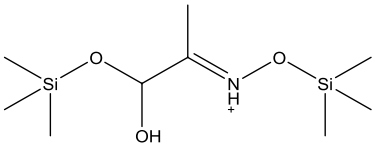
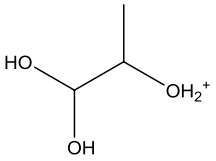
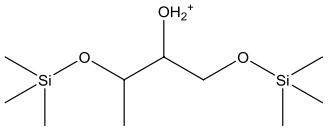
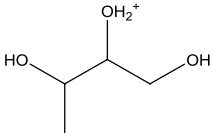
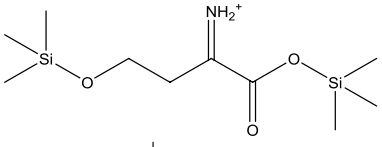
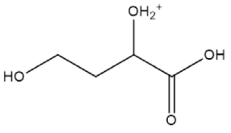
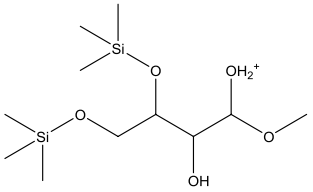
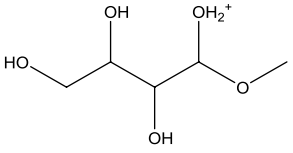
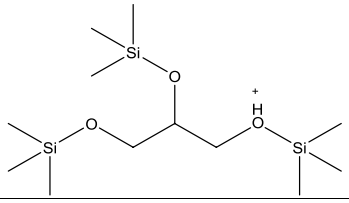
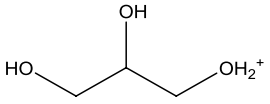
To confirm the structural differences among extracted pectins, an ANN has been trained on pectin composition parameters, leading to high prediction rates on new samples. This model could be applied on pectin extracted from different sources and samples from unknown origin to determine similar hydrolysis patterns. In addition, GC-MS spectra modelling of both identified and unidentified sugars detected in pectin hydrolysates found highly reproducible classification patterns. The study of relevant *m/z* ions suggest that several unidentified sugars

Table 4
Possible chemical structures of the most relevant *m/z* ions introduced as neural network inputs, determined by CFM-ID. These ions correspond to **TMSO fragments** from cell wall monosaccharides and cumaric and ferulic acids.

m/z	TMSO Structure	Original carbohydrate structure
107		
121		
133		
151		
193		
195		
197		
209		
234		

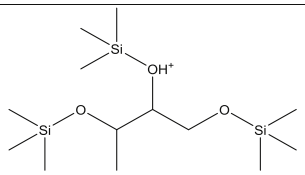
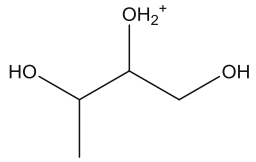
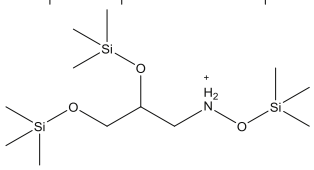
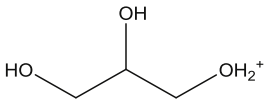
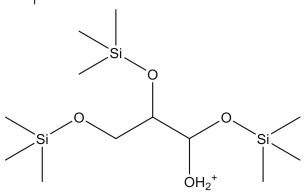
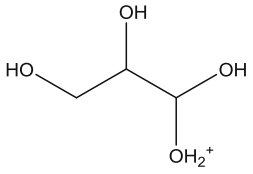
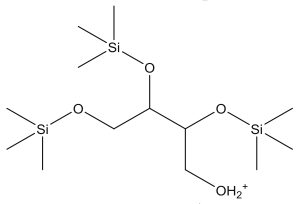
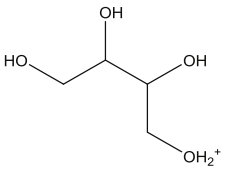
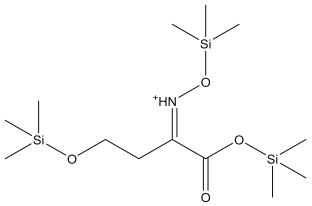
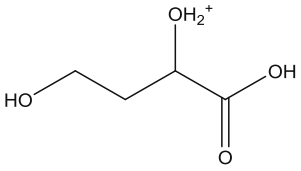
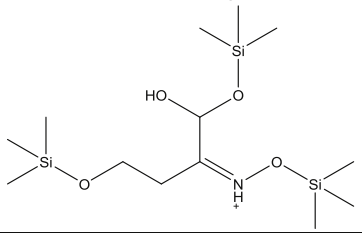
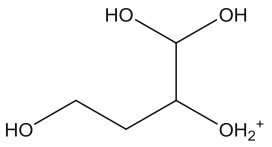
(*continued on next page*)

Table 4 (*continued*)

m/z	TMSO Structure	Original carbohydrate structure
235		
248		
250		
251		
262		
297		
309		

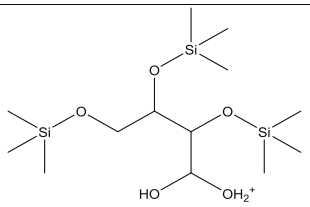
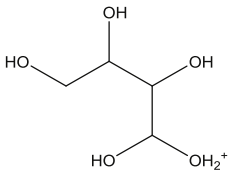
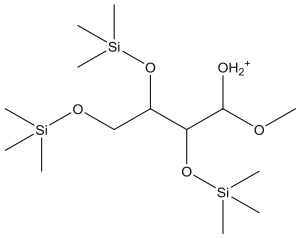
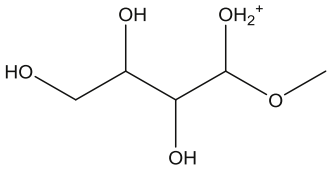
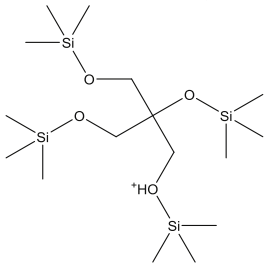
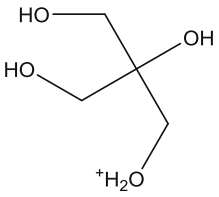
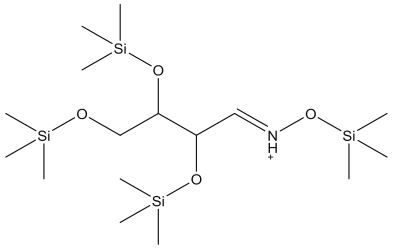
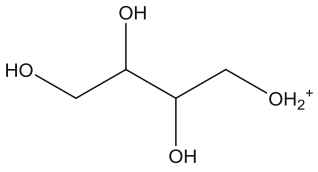
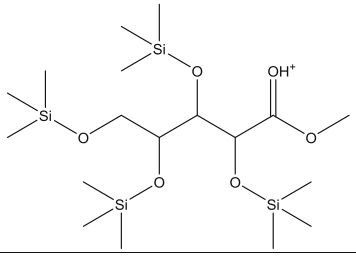
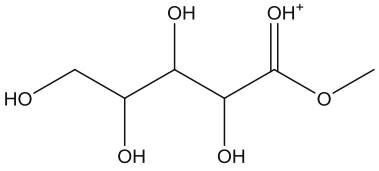
(*continued on next page*)

Table 4 (*continued*)

m/z	TMSO Structure	Original carbohydrate structure
323		
324		
325		
339		
350		
352		

(*continued on next page*)

Table 4 (*continued*)

m/z	TMSO Structure	Original carbohydrate structure
355		
369		
411		
424		
469		

may correspond to acidic derivatives of methyl-esterified/acetylated GalA.

Acknowledgments

This work has been funded by MICINN of Spain, Projects AGL2014-53445-R and AGL2017-84614-C2-1-R. Carlos Sabater thanks his FPU Predoc contract from Spanish MECD (FPU14/03619). Authors are also thankful to Riberebro (La Rioja, Spain) for kindly providing the artichoke by-products studied in this paper and Ramiro Martinez (Novozyme) for enzyme supply.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodhyd.2019.105238.

References

Adetunji, L. R., Adekunle, A., Orsat, V., & Raghavan, V. (2017). Advances in the pectin production process using novel extraction techniques: A review. *Food Hydrocolloids, 62*, 239–250.

Aldrich, E. (2013). *Wavelets: A package of functions for computing wavelet filters, wavelet transforms and multiresolution analyses. R package version 0.3-0.* URL https://CRAN.R-project.org/package=wavelets Last accessed: 24/06/2019 .

Allen, F., Greiner, R., & Wishart, D. (2015). Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification. *Metabolomics, 11*, 98–110.

Allen, F., Pon, A., Greiner, R., & Wishart, D. (2016). Computational prediction of electron ionization mass spectra to assist in GC/MS compound identification. *Analytical Chemistry, 88*(15), 7689–7697.

Atmodjo, M. A., Hao, Z., & Mohnen, D. (2013). Evolving views of pectin biosynthesis. *Annual Review of Plant Biology, 64*, 747–779.

Belscak-Cvitanovic, A., Dorcevic, V., Karlovic, S., Pavlovic, V., Komes, D., Jezek, D., et al. (2015). Protein-reinforced and chitosan-pectin coated alginate microparticles for delivery of flavan-3-ol antioxidants and caffeine from green tea extract. *Food Hydrocolloids, 51*, 361–374.

Bergmeir, C., & Benitez, J. M. (2012). Neural networks in R using the stuttgart neural network simulator: RSNNS. *Journal of Statistical Software, 46*, 1–26. URL http://www.jstatsoft.org/v46/i07/ Last accessed: 24/06/2019 .

Berlin, A., Maximenko, V., Gilkes, N., & Saddler, J. (2007). Optimization of enzyme complexes for lignocellulose hydrolysis. *Biotechnology and Bioengineering, 97*(2),

287–296.
Chandrapala, J., Oliver, C. M., Kentish, S., & Ashokkumar, M. (2013). Use of power ultrasound to improve extraction and modify phase transitions in food processing. *Food Reviews International, 29*(1), 67–91.
Delgado-Povedano, M. M., & de Castro, M. L. (2015). A review on enzyme and ultrasound: A controversial but fruitful relationship. *Analytica Chimica Acta, 889*, 1–21.
Dos Santos, E. K. R., Azoubel, P. M., & Gouveia, E. R. (2017). Better pectin yield from passion fruit peel (*Passiflora edulis* f. flavicarpa): From shaker or ultrasound? A comparison. *Waste and Biomass Valorization, 8*(3), 905–910.
Ebringerová, A., & Hromádková, Z. (2010). An overview on the application of ultrasound in extraction, separation and purification of plant polysaccharides. *Central European Journal of Chemistry, 8*(2), 243–257.
Fabris, A., Biasioli, F., Granitto, P. M., Aprea, E., Cappellin, L., Schuhfried, et al. (2010). PTR-TOF-MS and data-mining methods for rapid characterisation of agro-industrial samples: influence of milk storage conditions on the volatile compounds profile of Trentingrana cheese. *Journal of Mass Spectrometry, 45*(9), 1065–1074.
Fernández-Pierna, J. A., Volery, P., Besson, R., Baeten, V., & Dardenne, P. (2005). Classification of modified starches by Fourier transform infrared spectroscopy using support vector machines. *Journal of Agricultural and Food Chemistry, 53*(17), 6581–6585.
Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1–22. URL http://www.jstatsoft.org/v33/i01/ Last accessed: 24/06/2019 .
Huang, B., Zhao, K., Zhang, Z., Liu, F., Hu, H., & Pan, S. (2018). Changes on the rheological properties of pectin-enriched mango nectar by high intensity ultrasound. *LWT-Food Science and Technology, 91*, 414–422.
Jha, S. K., Josheski, F., Marina, N., & Hayashi, K. (2016). GC–MS characterization of body odour for identification using artificial neural network classifiers fusion. *International Journal of Mass Spectrometry, 406*, 35–47.
Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., & MacCoss, M. J. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature Methods, 4*, 923–925.
Lattanzio, V., & van Sumere, C. F. (1987). Changes in phenolic compounds during the development and cold storage of artichoke (*Cynara scolymus* L.) heads. *Food Chemistry, 24*(1), 37–50.
Li, X., He, X., Lv, Y., & He, Q. (2014). Extraction and functional properties of water--soluble dietary fiber from apple pomace. *Journal of Food Process Engineering, 37*(3), 293–298.
Lim, D. K., Long, N. P., Mo, C., Dong, Z., Cui, L., Kim, G., et al. (2017). Combination of mass spectrometry-based targeted lipidomics and supervised machine learning algorithms in detecting adulterated admixtures of white rice. *Food Research International, 100*, 814–821.
Lin, Z., Gonçalves, C. M. V., Dai, L., Lu, H. M., Huang, J. H., Ji, H., et al. (2014). Exploring metabolic syndrome serum profiling based on gas chromatography mass spectrometry and random forest models. *Analytica Chimica Acta, 827*, 22–27.
Liu, Y., Gong, G. L., Zhang, J., Jia, S. Y., Li, F., Wang, Y. Y., et al. (2014). Response surface optimization of ultrasound-assisted enzymatic extraction polysaccharides from *Lycium barbarum*. *Carbohydrate Polymers, 110*, 278–284.
Majka, M.. naivebayes: High performance implementation of the naive Bayes algorithm. (2019). R package version 0.9.5. URL https://CRAN.R-project.org/package=naivebayes Last accessed: 24/06/2019.
Marić, M., Grassino, A. N., Zhu, Z., Barba, F. J., Brnčić, M., & Brnčić, S. R. (2018). An overview of the traditional and innovative approaches for pectin extraction from plant food wastes and by-products: Ultrasound-, microwaves-, and enzyme-assisted extraction. *Trends in Food Science & Technology, 76*, 28–37.
Mohnen, D. (2008). Pectin structure and biosynthesis. *Current Opinion in Plant Biology, 11*(3), 266–277.
Morris, V. J., Belshaw, N. J., Waldron, K. W., & Maxwell, E. G. (2013). The bioactivity of modified pectin fragments. *Bioactive Carbohydrates and Dietary Fibre, 1*(1), 21–37.
Muñoz-Almagro, N., Montilla, A., Moreno, F. J., & Villamiel, M. (2017). Modification of citrus and apple pectin by power ultrasound: Effects of acid and enzymatic treatment. *Ultrasonics Sonochemistry, 38*, 807–819.
Muñoz-Almagro, N., Rico-Rodriguez, F., Wilde, P. J., Montilla, A., & Villamiel, M. (2018). Structural and technological characterisation of pectin extracted with sodium citrate and nitric acid from sunflower heads. *Electrophoresis, 39*, 1984–1992.
Nadar, S. S., Rao, P., & Rathod, V. K. (2018). Enzyme assisted extraction of biomolecules as an approach to novel extraction technology: A review. *Food Research International, 108*, 309–330.
Nguyen, T. T. T., & Le, V. V. M. (2013). Effects of ultrasound on cellulolytic activity of cellulase complex. *International Food Research Journal, 20*(2).
Ognyanov, M., Remoroza, C., Schols, H. A., Georgiev, Y., Kratchanova, M., & Kratchanov, C. (2016). Isolation and structure elucidation of pectic polysaccharide from rose hip fruits (*Rosa canina* L.). *Carbohydrate Polymers, 151*, 803–811.
Panchev, I. N., Kirtchev, N. A., & Kratchanov, C. G. (1994). On the production of low esterified pectins by acid maceration of pectic raw materials with ultrasound treatment. *Food Hydrocolloids, 8*(1), 9–17.
Peng, F., Peng, P., Xu, F., & Sun, R. C. (2012). Fractional purification and bioconversion of hemicelluloses. *Biotechnology Advances, 30*(4), 879–903.
Rico-Rodríguez, F., Serrato, J. C., Montilla, A., & Villamiel, M. (2018). Impact of ultrasound on galactooligosaccharides and gluconic acid production throughout a multienzymatic system. *Ultrasonics Sonochemistry, 44*, 177–183.
Ruttkies, C., Schymanski, E. L., Wolf, S., Hollender, J., & Neumann, S. (2016). MetFrag relaunched: Incorporating strategies beyond *in silico* fragmentation. *Journal of Cheminformatics, 8*, 3.
Sabater, C., Corzo, N., Olano, A., & Montilla, A. (2018a). Enzymatic extraction of pectin from artichoke (Cynara scolymus L.) by-products using Celluclast®1.5 L. *Carbohydrate Polymers, 190*, 43–49.
Sabater, C., Ferreira-Lazarte, A., Montilla, A., & Corzo, N. (Ferreira-Lazarte, Montilla, & Corzo, 2019a). Enzymatic production and characterization of pectic oligosaccharides derived from citrus and apple pectins: A GC-MS study using random forests and association rule learning. *Journal of Agricultural and Food Chemistry, Article ASAP*. https://doi.org/10.1021/acs.jafc.9b00930.
Sabater, C., Montilla, A., Ovejero, A., Prodanov, M., Olano, A., & Corzo, N. (2018b). Furosine and HMF determination in prebiotic-supplemented infant formula from Spanish market. *Journal of Food Composition and Analysis, 66*, 65–73.
Sabater, C., Olano, A., Corzo, N., & Montilla, A. (Olano, Corzo, & Montilla, 2019b). GC–MS characterisation of novel artichoke (*Cynara scolymus*) pectic-oligosaccharides mixtures by the application of machine learning algorithms and competitive fragmentation modelling. *Carbohydrate Polymers, 205*, 513–523.
Soria, A. C., Villamiel, M., & Montilla, A. (2017). *Ultrasound effects on processes and reactions involving carbohydrates. Ultrasound in food processing: Recent advances.* UK: John Wiley & Sons, Ltd437–463 Chichester.
Szabó, O. E., & Csiszár, E. (2013). The effect of low-frequency ultrasound on the activity and efficiency of a commercial cellulase enzyme. *Carbohydrate Polymers, 98*(2), 1483–1489.
Tao, J. X., Huang, J. B., Yu, L., Li, Z. K., Liu, H. S., Yuan, B., et al. (2018). A new methodology combining microscopy observation with Artificial Neural Networks for the study of starch gelatinization. *Food Hydrocolloids, 74*, 151–158.
Tao, Y., Wu, D., Zhang, Q. A., & Sun, D. W. (2014). Ultrasound-assisted extraction of phenolics from wine lees: Modeling, optimization and stability of extracts during storage. *Ultrasonics Sonochemistry, 21*(2), 706–715.
Uarrota, V. G., Moresco, R., Coelho, B., da Costa Nunes, E., Peruch, L. A. M., de Oliveira Neubert, E., et al. (2014). Metabolomics combined with chemometric tools (PCA, HCA, PLS-DA and SVM) for screening cassava (*Manihot esculenta* Crantz) roots during postharvest physiological deterioration. *Food Chemistry, 161*, 67–78.
Wang, W., Ma, X., Jiang, P., Hu, L., Zhi, Z., Chen, J., et al. (2016). Characterization of pectin from grapefruit peel: A comparison of ultrasound-assisted and conventional heating extractions. *Food Hydrocolloids, 61*, 730–739.
Wang, W., Ma, X., Xu, Y., Cao, Y., Jiang, Z., Ding, T., et al. (2015). Ultrasound-assisted heating extraction of pectin from grapefruit peel: Optimization and comparison with the conventional method. *Food Chemistry, 178*, 106–114.
Wang, J., Sun, B., Liu, Y., & Zhang, H. (2014). Optimisation of ultrasound-assisted enzymatic extraction of arabinoxylan from wheat bran. *Food Chemistry, 150*, 482–488.
Wang, W., Wu, X., Chantapakul, T., Wang, D., Zhang, S., Ma, X., et al. (2017). Acoustic cavitation assisted extraction of pectin from waste grapefruit peels: A green two-stage approach and its general mechanism. *Food Research International, 102*, 101–110.
Wikiera, A., Mika, M., Starzyńska-Janiszewska, A., & Stodolak, B. (2015). Development of complete hydrolysis of pectins from apple pomace. *Food Chemistry, 172*, 675–680.
Xu, J. K., Li, M. F., & Sun, R. C. (2015). Identifying the impact of ultrasound-assisted extraction on polysaccharides and natural antioxidants from *Eucommia ulmoides* Oliver. *Process Biochemistry, 50*(3), 473–481.
Yang, Y., Wang, Z., Hu, D., Xiao, K., & Wu, J. Y. (2018). Efficient extraction of pectin from sisal waste by combined enzymatic and ultrasonic process. *Food Hydrocolloids, 79*, 189–196.
Yapo, B. M., & Koffi, K. L. (2008). Dietary fiber components in yellow passion fruit rind: A potential fiber source. *Journal of Agricultural and Food Chemistry, 56*(14), 5880–5883.
Yuliarti, O., Goh, K. K., Matia-Merino, L., Mawson, J., & Brennan, C. (2015). Extraction and characterisation of pomace pectin from gold kiwifruit (*Actinidia chinensis*). *Food Chemistry, 187*, 290–296.

Carbohydrate Polymers 205 (2019) 513–523

Contents lists available at ScienceDirect

Carbohydrate Polymers

journal homepage: www.elsevier.com/locate/carbpol

GC–MS characterisation of novel artichoke (*Cynara scolymus*) pectic-oligosaccharides mixtures by the application of machine learning algorithms and competitive fragmentation modelling

Carlos Sabater, Agustín Olano, Nieves Corzo*, Antonia Montilla

Instituto de Investigación en Ciencias de la Alimentación CIAL, (CSIC-UAM) CEI (UAM + CSIC), C/Nicolás Cabrera, 9, E-28049, Madrid, Spain

ARTICLE INFO

Keywords:
Artichoke pectin
Enzymatic hydrolysis
Pectic-oligosaccharides
Neural network
In silico fragmentation

ABSTRACT

Novel artichoke pectic-oligosaccharides (POS) mixtures have been obtained by enzymatic hydrolysis using four commercial enzyme preparations: Glucanex®200G, Pentopan®Mono-BG, Pectinex®Ultra-Olio and Cellulase from *Aspergillus niger*. Analysis by HPAEC-PAD showed that Cellulase from *A. niger* produced the greatest amount of POS (310.6 mg g^{-1} pectin), while the lowest amount was produced by Pentopan®Mono-BG (45.7 mg g^{-1} pectin). To determine structural differences depending on the origin of the enzyme, GC–MS spectra of di- and tri-saccharides have been studied employing three machine learning algorithms: multilayer perceptron, random forest and boosted logistic regression. Machine learning models allowed characteristic *m/z* ions patterns to be established for each enzyme based on their GC–MS spectra with high prediction rates (above 95% on the test set). Possible chemical structures were given for some *m/z* ions having a decisive influence on these classifications. Finally, it was observed that several ions could be formed from specific POS structures.

1. Introduction

Pectin, one of the most structurally complex families of polysaccharides in nature, is mainly composed of linear chains of α-1,4-D-galacturonic acid (GalA) called homogalacturonan (HG), which comprise approximately 70% of total pectin. The other two domains present ramified chains as rhamnogalacturonan type I (RG-I), a polymer with alternate sequences of GalA and α-(1, 2) linked L-rhamnosyl residues which may be substituted at *O*-4 with linear or branched oligosaccharides, and rhamnogalacturonan II (RG-II), a most complex structure consisting of 12 different types of sugars in over 20 different linkages (Gullón et al., 2013; Mohnen, 2008). Pectins present interesting properties which are directly influenced by its structural characteristics (monomeric composition, presence and distribution of side chains, degree of methyl-esterification and acetylation, molar mass, and charge distribution along their backbone) (Herbstreith & Fox, 2018; Maric et al., 2018).

The most commonly used pectins in the food industry are extracted from citrus peel, apple and sugar-beet by-products. However, the recent optimization of a procedure of pectin extraction from artichoke by-products could provide a new source with interesting techno-functional properties (Sabater, Corzo, Olano, & Montilla, 2018).

By the partial depolymerization of pectins through chemical or enzymatic methods pectic-oligosaccharides (POS) can be obtained and it is known that POS are one of the most promising candidates to be recognised as prebiotics (Gullón et al., 2013). Enzymatic hydrolysis of pectins shows several advantages, such as high regio- and stereo-selectivity, obtaining structurally different POS mixtures (Babbar, Dejonghe, Gatti, Sforza, & Elst, 2016; da Moura, Macagnan, & da Silva, 2015). Most food-grade POS studied are produced by commercial pectinases with three different enzymatic activities polygalacturonase (PG), pectin lyase (PL) and pectin esterase (PE) involving reactions of hydrolysis, β-elimination and removal of methyl or acetyl groups, respectively. Commercial pectinases are usually produced by *Aspergillus* spp. (Combo, Aguedo, Goffin, Wathelet, & Paquot, 2012).

POS mixtures can be formed of various types of oligomers and these structures are often characterised using chromatographic techniques such as HPAEC-PAD (Babbar, Dejonghe, Sforza, & Elst, 2017; Combo et al., 2013; Gómez, Gullón, Yáñez, Parajó, & Alonso, 2013) and mass spectrometry (MS) (Leijdekkers, Huang, Bakx, Gruppen, & Schols, 2015; Ognyanov et al., 2016). MS allows structural patterns of complex molecules to be determined and has become the analytical method of choice in metabolomics research. However, it generates large amounts of intricate data that needs to be interpreted to extract chemical information and ensure that they are of valuable (Boccard & Rudaz, 2014; Yi et al., 2016). The identification of unknown compounds is the main

* Corresponding author.
E-mail address: nieves.corzo@csic.es (N. Corzo).

https://doi.org/10.1016/j.carbpol.2018.10.054
Received 3 August 2018; Received in revised form 11 October 2018; Accepted 18 October 2018
Available online 24 October 2018
0144-8617/

bottleneck, so computational tools assisting structure elucidation and *de novo* identification of small molecules have been developed including *in silico* fragmentation methods (Allard et al., 2016; Ruttkies, Schymanski, Wolf, Hollender, & Neumann, 2016). These methods help in the deduction of possible structures of metabolites in spectra interpretation. *Competitive fragmentation modelling* (CFM) has been proposed as an *in silico* fragmentation method suitable for common MS techniques such as GC-MS-EI and significantly outperforms existing state-of-the-art computational methods (Allen, Greiner, & Wishart, 2015; Allen, Pon, Greiner, & Wishart, 2016). It has been validated and tested on the NIST database, and it could be used as an alternative when no reference standard is available for measurement.

Other computational tools are focused on data modelling to find reproducible patterns and discover valuable information on biological events and chemical/structural properties (Käll, Canterbury, Weston, Noble, & MacCoss, 2007; Yi et al., 2016). A number of machine learning classification methods have been applied in the analysis of MS spectra. Supervised and semi-supervised classification methods support *a priori* known data structures to train patterns and rules to predict new data. When the relationship between MS data and chemical structures is complex, simple classifiers may not be efficient. Today there exists a multitude of algorithms to determine classification patterns among samples, including support vector machines (SVM) and random forests (RF), which can also be applied to the field of metabolomics (Lin et al., 2014; Uarrota et al., 2014) and also food science (Lim et al., 2017; Sabater, Montilla et al., 2018). Tree-based models have been used for proteomic mass spectra classification (Geurts et al., 2005) and recent works reported high prediction rates based on MALDI-TOF data (Rossel & Martínez Arbizu, 2018). Artificial neural networks (ANN) and boosted models are other machine learning algorithms used to classify spectral data of small molecules (Gosav, Praisler, & Birsa, 2011), which recognize chemical substructures from MS data (Varmuza, He, & Fang, 2003), feature selection in MS-based proteomic profiling (Gertheiss & Tutz, 2009) or protein biomarker discovery (Yasui et al., 2003). It has been reported that machine learning has greatly improved performance relative to the bond-breaking approaches and even CFM (Schymanski et al., 2017).

Therefore, the aim of this study was characterised by GC–MS novel artichoke POS obtained by enzymatic hydrolysis using commercial enzyme preparations with different main activities. In order to accomplish this, structural characteristics of di- and tri-POS were determined by mass spectral mining using three machine learning algorithms: multilayer perceptron (MLP), random forest (RF) and boosted logistic regression (BLR).

2. Materials and methods

2.1. Standards and reagents

Analytical reference substances such as sucrose, D-arabinose, D-xylose, L-rhamnose, D-galactose, D-mannose, D-glucose, galacturonic acid (GalA), digalacturonic acid (Di-GalA), kestose, nystose and β-phenyl glucoside were purchased from Sigma Aldrich (Steinheim, Germany). Trigalacturonic acid (Tri-GalA) was from Carbosynth (Compton, UK). Four commercial enzyme preparations were studied (Table 1). Cellulase from *Aspergillus niger* was acquired from Sigma Aldrich (Steinheim, Germany). The rest of the enzyme preparations were a generous gift from Novozymes (Bagsvaerd, Denmark).

Artichoke pectin was previously extracted in our laboratory using a commercial cellulase, Celluclast®1.5 L (artichoke by-product powder concentration 6.5%, enzyme dose 10.1 U g^{-1}, extraction time 48 h). This pectin has GalA content of 69.5%, degree of methylation of 19.5% and molecular mass (Mw) ranging from 4.8 to 660 kDa (Sabater, Corzo et al., 2018).

2.2. Enzyme characterisation

Enzyme characterisation assays were carried out following the method described by Martínez, Gullón, Yanez, Alonso, and Parajó (2009) with modifications. A solution of 2% (w/v) of polygalacturonic acid, chosen as HG standard (Sigma, purity > 85%, GalA content greater than 96%), was dissolved in 0.05 M acetate buffer (pH 5.0). Pectinase activity was measured by quantifying the amount of reducing sugar groups liberated after incubation of polygalacturonic acid solutions with 5 mg mL^{-1} or 5 μL mL^{-1} of enzyme at 50 °C for 5 min, using the method of DNS and GalA as standard. One unit of pectinase activity was defined as the amount of enzyme required to release 1 μmol of GalA per min at 50 °C.

2.3. Formation of pectic-oligosaccharides (POS)

Pectic-oligosaccharides (POS) were obtained by enzymatic hydrolysis of 2% (w/v) artichoke pectin solutions dissolved in 0.05 M acetate buffer (pH 5.0) using 0.54–6.75 U mL^{-1} of enzyme (Table 1) following the method of Gómez, Yañez, Parajo, and Alonso (2016) with some modifications.

Enzymatic hydrolysis was performed in a final volume of 1 mL incubated in an orbital shaker at 50 °C and 750 rpm. Aliquots were withdrawn from the reaction mixture at the different times (0.5, 1 and 4 h) and immediately immersed in boiling water for 5 min to inactivate the enzyme. Samples were stored at −18 °C for subsequent analysis. Enzymatic reactions were carried out in duplicate and analyses were performed twice for each enzymatic treatment. In addition, four hydrolysis replicates were prepared for reactions incubated at 4 h for their GC–MS characterization.

2.4. Analytical techniques

Monosaccharides were quantified by GC-FID as trimethyl silylated oximes (TMSO) (Sabater, Corzo et al., 2018). In addition, samples containing di- and tri-POS formed were analysed by GC–MS on an Agilent Technologies 7890 A gas chromatograph coupled to a 5975CMSD quadrupole mass detector (Agilent Technologies) to characterise low M_w POS obtained. Separations were carried out using a fused silica capillary column HP-5MS (5% phenyl methylsilicone, 30 m × 0.25 mm × 0.25 μm thickness; J&W Scientific, Folsom, CA,

Table 1
Determination of commercial enzyme preparation activities (U g^{-1} and U mL^{-1}) using polygalacturonic acid as substrate.

Enzyme preparation	Microorganism	Declared activity	Hydrolase activity (U g^{-1})	Enzyme dose** (U mL^{-1})
Pectinex®Ultra-Olio	*Aspergillus aculeatus/ Aspergillus niger*	Pectin-lyase	202.6 ± 7.4[a,b,*]	6.75
Glucanex®200 G	*Trichoderma harzianum*	Glucanase	189.2 ± 28.5[b]	0.63
Pentopan®Mono-BG	*Thermomyces lanuginosus*	1,4-endoxylanase	161.5 ± 3.6[b]	0.54
Cellulase from *A. niger*	*Aspergillus niger*	Cellulase	263.7 ± 18.5[a]	

[a,b]Statistically significant differences between enzymes.
* Enzyme preparation in liquid form: U mL^{-1}.
** Enzyme dose used for POS production.

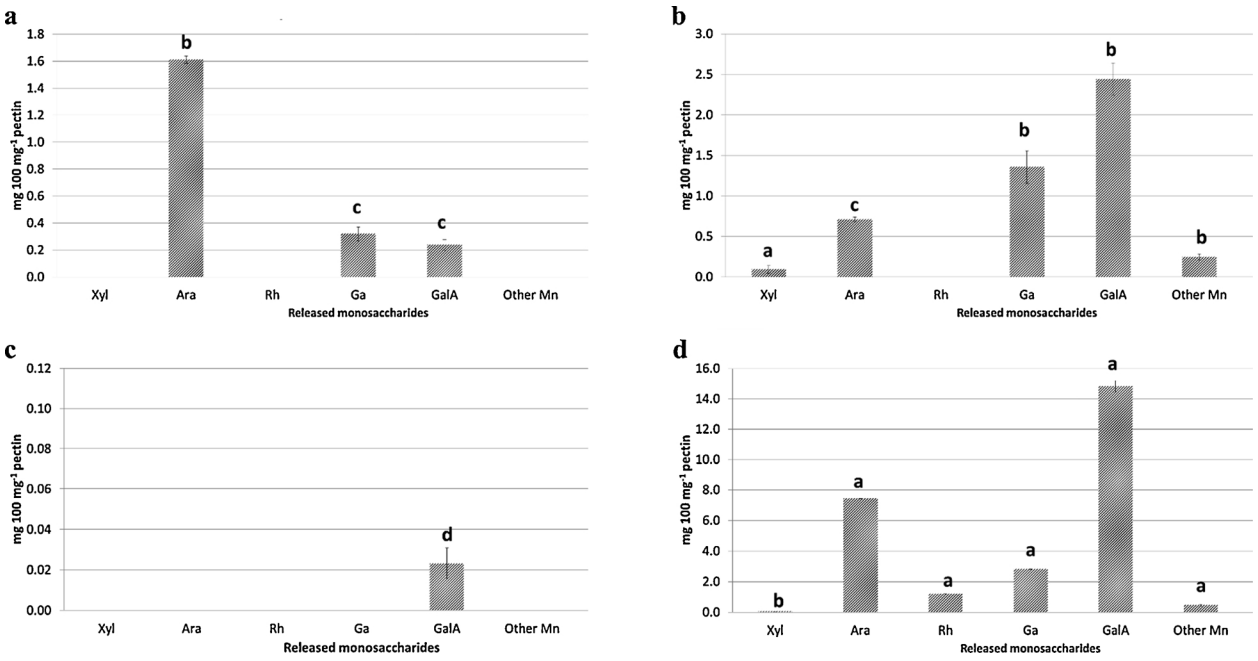


Fig. 1. Monosaccharides (mg 100 mg^{-1} pectin) found in enzymatic hydrolysates of artichoke pectin after 4 h of reaction using: a) Glucanex®200 G (0.63 U mL^{-1}); b) Cellulase from *Aspergilus niger* (0.88 U mL^{-1}); c) Pentopan®Mono-BG (0.54 U mL^{-1}); d) Pectinex®Ultra-Olio (6.75 U mL^{-1}). Xyl: xylose, Ara: arabinose, Rh: rhamnose, Ga: galactose, GalA: galacturonic acid, Other Mn: unidentified monosaccharides. [a,b,c,d] Statistically significant differences between enzymes.

USA). Helium was used as carrier gas at a flow rate of 0.8 mL min^{-1}. Injector temperature was 200 °C. The oven temperature was initially 200 °C and held for 5 min, then increased at a rate of 3 °C min^{-1} to 250 °C and held for 1 min, then increased at a rate of 10 °C min^{-1} to 320 °C and held for 70 min. Injections were made in the split mode (1:5). The mass spectrometer was operated in electrospray ionisation mode at 70 eV. Mass spectra were acquired using Agilent ChemStation MSD software. Internal standard (β-phenyl glucoside) was added to the samples. Identification of trimethylsilyl oxime derivatives of carbohydrates was carried out by comparison of their relative retention times and mass spectra with those of standard compounds (GalA, Di-GalA and Tri-GalA).

Average M_w distribution of POS formed were determined by HPSEC-ELSD following the methods described by Sabater, Corzo et al. (2018). For HPSEC-ELSD analysis, samples (0.65 mg mL^{-1}) were dissolved in water filtered using a 0.45 µm syringe filter (Symta, Spain) and analysed in an Agilent Technologies 1220 Infinity LC System (Böblingen, Germany) The separation of carbohydrates was carried out on a TSK-GEL G5000PW$_{XL}$ column (300 mm x 7.8 mm, 10 µm particle size) and TSK-GEL G2500PW$_{XL}$ column (300 mm x 7.8 mm, 6 µm particle size) in combination with a TSK-GEL PW$_{XL}$ guard column (40 mm x 6 mm, 12 µm particle size) (Tosoh Bioscience, Montgomeryville, PA, USA) using 0.01 M NH_4Ac, as mobile phase and elution in isocratic mode at a flow rate of 0.5 mL min^{-1} for 50 min. Molecular weight of carbohydrates was calculated by the external calibration method using solutions of commercial pullulan standards (M_w 0.342–788 kDa) (Fluka Analytical) in the range 10–2250 mg L^{-1}.

POS obtained with different commercial enzyme preparations were quantified by HPAEC-PAD using a DIONEX ICS2500 system (Dionex Corp., Sunnyvale, CA, USA) incorporating a GP50 gradient pump and an ED50 electrochemical detector using a gold and Ag/AgCl as working and reference electrodes, respectively. Separations were carried out at room temperature in a CarboPac PA-1 column (2 × 450 mm) and a CarboPac PA 1 guard column (4 × 50 mm) as a flow rate of 1 mL min^{-1}. The mobile phases were (A) 0.1 M NaOH and (B) 1 M NaOAc in 0.1 M NaOH). The elution profile was as follows: 0-15 min, 0–5% B; 15–60 min, 5–70% B; 60–65 min, 70–100% B; 65–70 min, 100% B; 70–70.1 min, 100-0% B; and finally column re-equilibration by 0% B from 70.1 to 85 min. Pectin neutral monosaccharides (arabinose, xylose, rhamnose, galactose), neutral di- tri- and tetrasaccharides (sucrose, kestose and nystose) as well as GalA and its derivatives (Di-GalA and Tri-GalA) standards were used for identification. Acquisition and processing of data were achieved with Chromeleon 6.7 software (Dionex Corp. CA, USA).

2.5. Data analysis

ANOVA tests and Tukey´s test for $p < 0.05$ were applied to all data generated.

Structural characteristics of POS obtained with Pentopan®Mono-BG, Glucanex®200 G, Cellulase from *A. niger* and Pectinex®Ultra-Olio were determined by GC−MS spectral mining. In order to find reproducible patterns which could be applied on new samples, three machine learning models were evaluated in a supervised classification task: multilayer perceptron (MLP), random forest (RF) and boosted logistic regression (BLR). To ensure these algorithms were trained with valuable information, *m/z* fragments whose abundances were statistically different among groups ($p < 0.05$) corresponding to known POS ruptures were selected.

2.5.1. Signal processing

To increase model performance, discrete wavelet transform (DWT) was applied to GC–MS data (Li, Li, & Yao, 2007; Xia, Wu, & Yuan, 2007). In this study, a 16 level DWT was computed (indicating the depth of the decomposition) with a “la8” decomposition filter (Daubechies orthonormal compactly supported wavelet of length = 8). In the threshold step, DWT coefficients under percentile 15% were removed to denoise the signal. Once the signal was reconstructed, all variables were scaled and centered.

2.5.2. Classification of MS spectra

462 mass spectra of known pectic sugars and unknown POS released during enzymatic hydrolysis were classified using MLP, RF and BLR.

MLP is the most common kind of artificial neural network (ANN), a family of broadly used models that allow modeling complex and highly non-linear processes. MLP is formed by an input layer (i.e. pre-

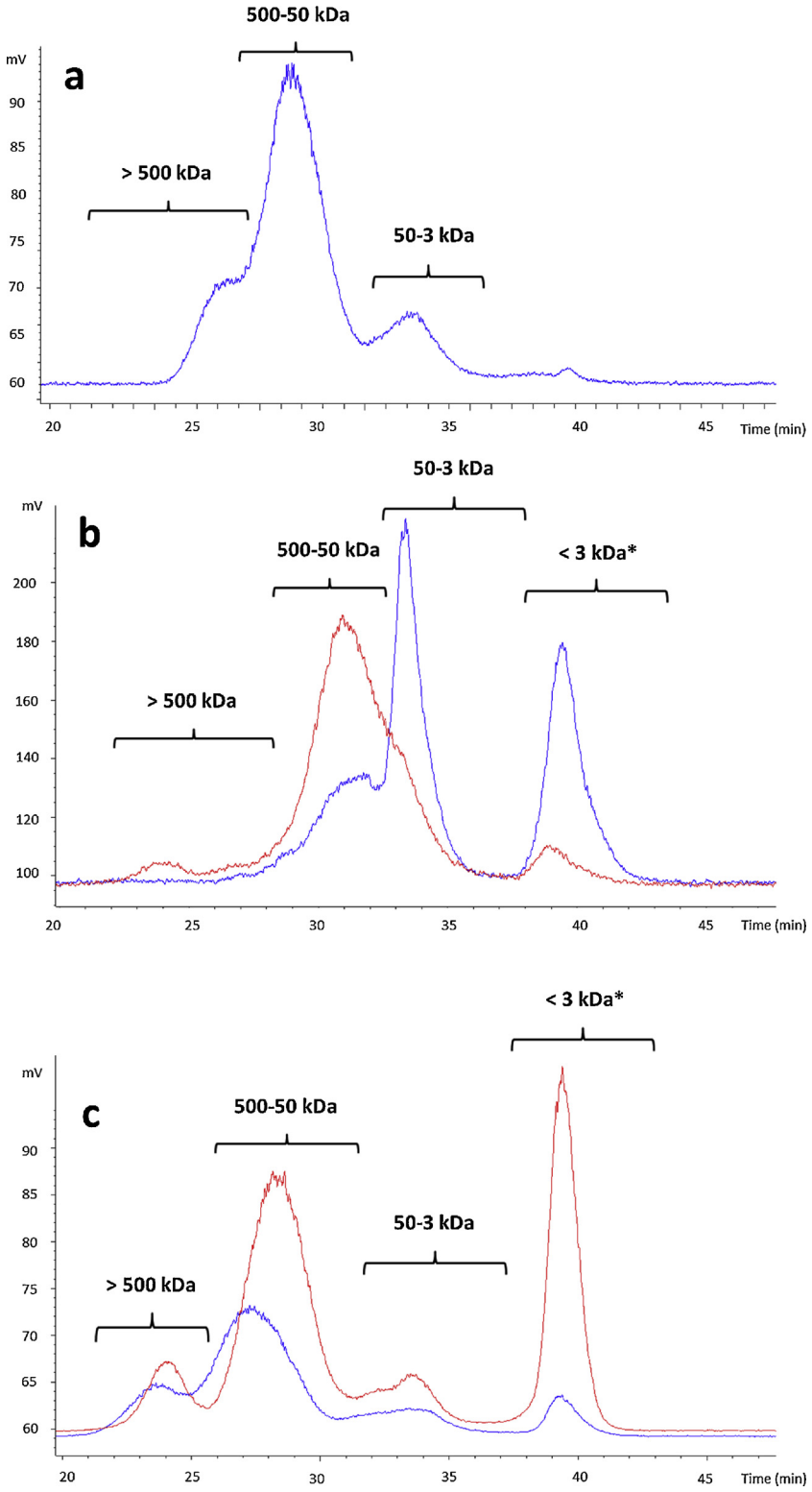


Fig. 2. HPSEC-ELSD profiles and molecular weight ranges of a) artichoke pectin; and hydrolysates obtained from artichoke pectin after 4 h of reaction using b) Cellulase from *A. niger* (red) and Pectinex®Ultra-Olio (blue), c) Pentopan®Mono-BG (blue) and Glucanex®200 G (red). *Compounds not quantified due to coelution with other present in enzymatic preparations (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

processed GC–MS spectra), an output layer (i.e. enzyme used to obtain POS) and several neurons or nodes organised in hidden layers, where each neuron in a layer is connected with each neuron in the next layer through a weighted connection. In this case, an MLP was built with 1 hidden layer consisting in 25 neurons. The activation function (a transformation applied to the input signal to determine whether the information that the neuron is receiving is relevant or not) was logistic.

In RF, a multitude of decision trees are constructed, outputting the different classes. Each node is split using the best among a subset of predictors randomly chosen. In this case, RF model was built with 500

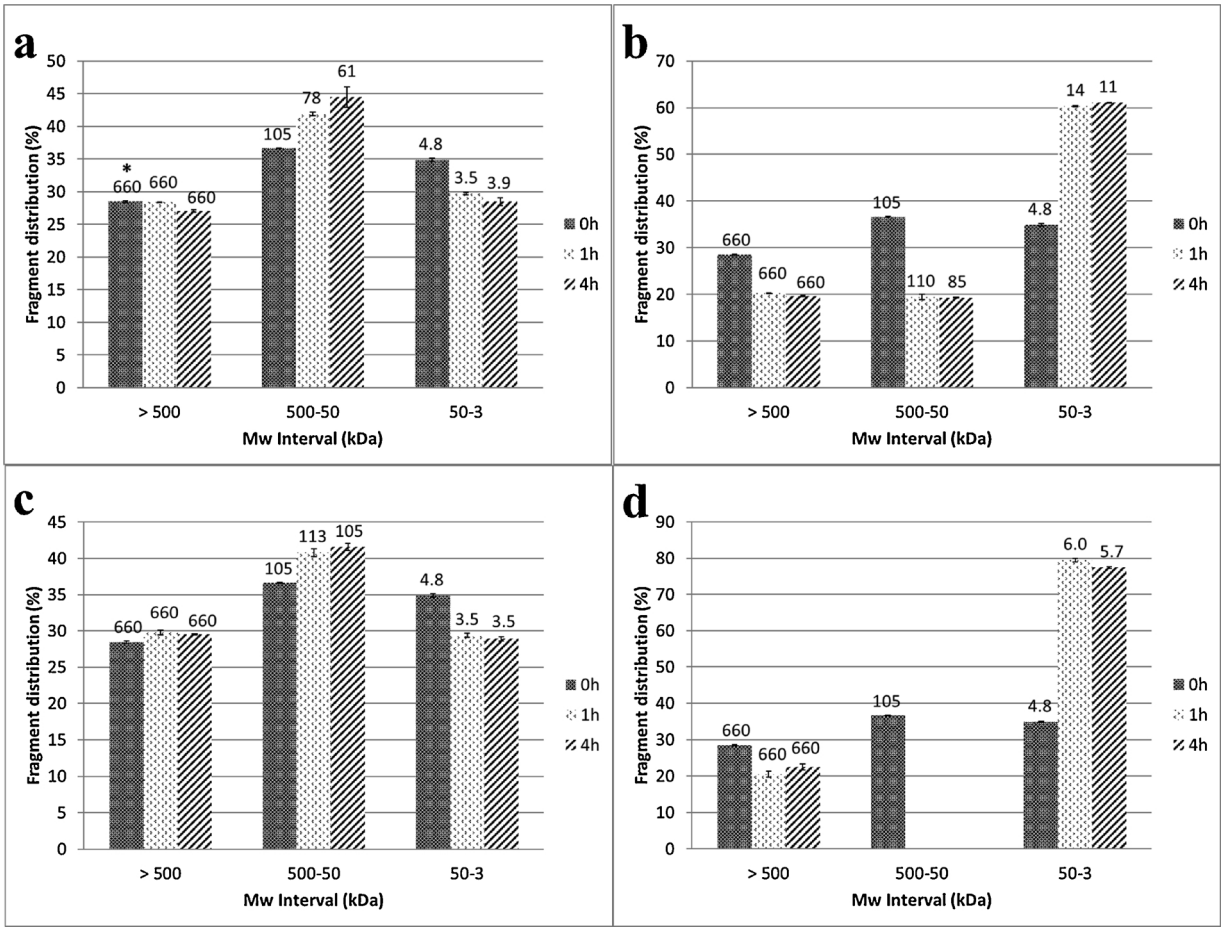


Fig. 3. Molecular mass distribution of fragments with M_w > 3 kDa determined by HPSEC-ELSD of different hydrolysates of artichoke pectin after 1 and 4 h of reaction using: a) Glucanex®200 G (0.63 U mL^{-1}); b) Cellulase from *Aspergilus niger* (0.88 U mL^{-1}); c) Pentopan®Mono-BG (0.54 U mL^{-1}); d) Pectinex®Ultra-Olio (6.75 UmL^{-1}). *Average M_w of the fragments within M_w interval.

trees and 50 variables tried at each split.

On the other hand, BLR is considered as an ensemble method that uses a weighted average of predictions of individual classifiers (Geurts et al., 2005). The iterations specify the maximal number of trees to be fitted. In this work, a BLR consisting in 11 iterations was chosen.

All the models were trained with 70% of the data, 10-fold cross-validated and then tested with 30% of data from each group (corresponding to new samples). Variable importance analysis was carried out to determine the most influential *m/z* ions in each model. In MLP, influent *m/z* ions were determined by the sum of the product of raw input-hidden and hidden-output connection weights while a permutation of the out of-bag-error was chosen for RF (an estimation of the classification error as trees are added to the forest). For BLR, importance coefficients were determined by calculating the area under the ROC (Receiver Operating Characteristic) curve.

All statistical analyses were computed on R v3.5.0. DWT was performed using wavelets package (Aldrich, 2013). MLP was computed using the RSNNS package (Bergmeir & Benitez, 2012) and RF classification was performed with Random Forest package (Liaw & Wiener, 2002). For BLR, caTools package was employed (Tuszynski, 2014).

2.5.3. *In silico fragmentation*

After determining the most important *m/z* ions in di- and tri-POS classification, chemical structures of some fragments have been proposed employing the competitive fragmentation modeling source code (CFM-ID) developed by Allen et al. (2016). GC-MS-EI fragmentation patterns of POS structures described in the bibliography (Atmodjo, Hao, & Mohnen, 2013) were determined and compared to those of enzymatically obtained POS.

3. Results and discussion

The polygalacturonase activity of four commercial enzyme preparations using polygalacturonic acid as a substrate was studied. As can be seen in Table 1, Pectinex Olio, Cellulase from *A. niger*, Glucanex and Pentopan showed high activities (202.6 U mL^{-1}, 263.7, 189.2 and 161.5 U g^{-1}, respectively. Moreover, these enzymes have different declared enzymatic activities including cellulase, pectin-lyase, exo-β-D-galactofuranosidase and 1, 4-endoxylanase, respectively and, therefore, different hydrolysis patterns could be expected. The complementary activities could be of great importance, so Cellulase from *A. niger* presented higher polygalacturonase activity than cellulase activity using polygalacturonic acid and carboxymethyl cellulose as substrates (Sabater, Corzo et al., 2018).

3.1. *Pectic-oligosaccharides (POS) obtainment from artichoke pectin hydrolysis*

Enzymatic obtainment of POS derived from artichoke pectin using the four tested enzymes was studied. In Table 1 the dose of each enzyme used to hydrolyse artichoke pectin is shown. Monosaccharides released during enzymatic hydrolysis were quantified by GC-FID. Fig. 1 shows the amounts found of each compound, GalA, neutral monosaccharides derived from pectins, such as xylose, arabinose, rhamnose and galactose as well as other unidentified monosaccharides present in artichoke hydrolysates. Pectinex Olio released significantly higher amounts of

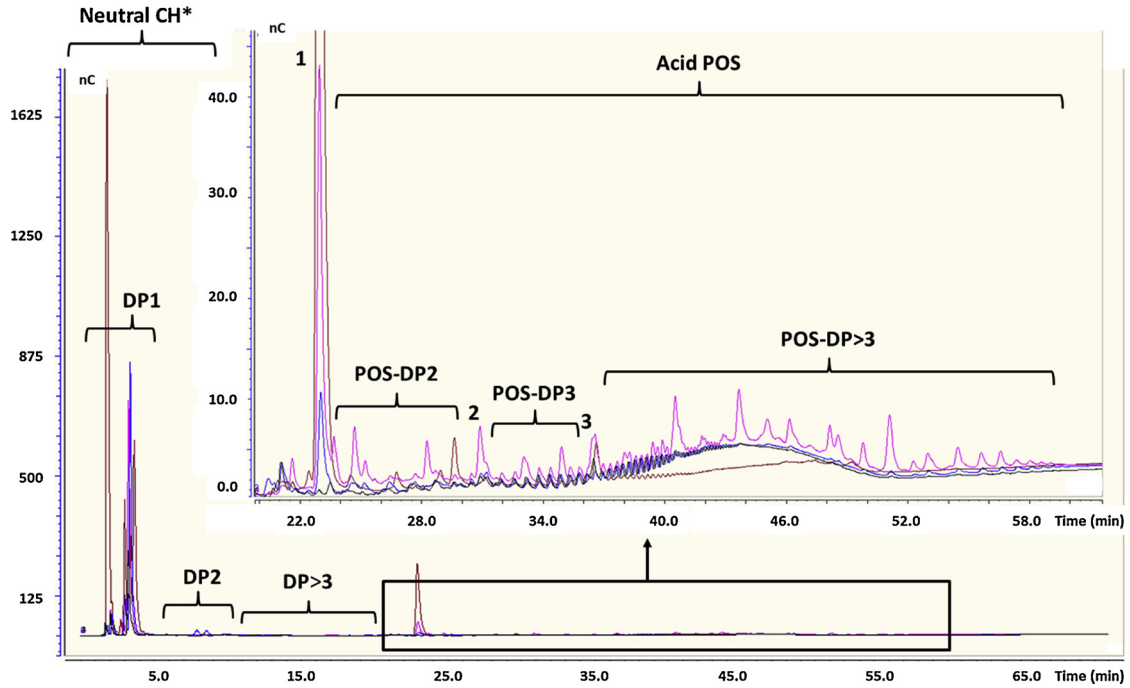


Fig. 4. HPAEC-PAD profiles of carbohydrates found in artichoke pectin hydrolysates using commercial enzyme preparation Cellulase from *A. niger* (pink), Pectinex Olio (brown), Glucanex (blue) and Pentopan (black) (50 °C, pH = 5.0, 4 h, 0.54–6.75 U mL^{-1} enzyme). CH: carbohydrates, POS: pectic-oligosaccharides. Peaks: DP1, DP2 and DP3 neutral mono-, di- and oligosaccharides DP > 3**,** neutral oligosaccharides; (1) Galacturonic acid; (2) Digalacturonic acid; (3) Trigalacturonic acid; POS-DP2, acid disaccharides; POS-DP3, acid trisaccharides; POS-DP > 3, others acid oligosaccharides. (*) Compounds not quantified (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Table 2
POS formation (HPAEC-PAD) and galacturonic acid (GalA) release (mg g^{-1} pectin) during enzymatic hydrolysis (0.5, 1 and 4 h) of artichoke pectin using the four studied enzymes. Di-GalA: digalacturonic acid, Tri-GalA: trigalacturonic acid, POS: pectic-oligosaccharides.

Hydrolysis	Time (h)	Acid POS (mg g^{-1} pectin)						
		GalA (peak 1)	POS-DP2	Di-GalA (peak 2)	POS-DP3	Tri-GalA (peak 3)	POS-DP > 3	Total POS*
Pentopan®Mono-BG	0.5	3.0 ± 0.2 ^{d}	12.2 ± 1.0 ^{d}	7.8 ± 0.6 ^{b}	6.4 ± 0.2 $^{c, d, e}$	7.0 ± 0.5 $^{a, b}$	–	33.4 ± 2.8 ^{e}
	1	6.5 ± 0.6 ^{d}	12.6 ± 0.9 ^{d}	8.0 ± 0.6 ^{b}	7.7 ± 0.5 $^{c, d}$	7.4 ± 0.2 $^{a, b}$	–	35.7 ± 2.5 ^{e}
	4	10.0 ± 0.7 ^{d}	12.8 ± 1.1 ^{d}	8.2 ± 0.6 ^{b}	7.3 ± 0.2 $^{c, d}$	7.4 ± 0.5 $^{a, b}$	–	45.7 ± 3.9 ^{e}
Glucanex®200 G	0.5	6.5 ± 0.6 ^{d}	17.9 ± 1.3 $^{b, c, d}$	7.3 ± 0.5 ^{b}	0.1 ± 0.0 ^{e}	6.2 ± 0.2 $^{b, c}$	9.6 ± 0.8 ^{c}	41.1 ± 2.9 ^{e}
	1	11.0 ± 0.8 ^{d}	20.2 ± 1.4 $^{b, c}$	7.3 ± 0.5 ^{b}	4.6 ± 0.1 $^{d, e}$	7.7 ± 0.5 $^{a, b}$	22.1 ± 1.6 ^{c}	61.9 ± 5.3 $^{d, e}$
	4	11.4 ± 1.0 ^{d}	23.3 ± 1.6 $^{a, b}$	7.7 ± 0.5 ^{b}	6.9 ± 0.5 $^{c, d, e}$	7.1 ± 0.2 $^{a, b}$	18.7 ± 1.6 ^{c}	63.7 ± 4.5 $^{d, e}$
Cellulase from *A. niger*	0.5	15.0 ± 1.1 ^{d}	17.9 ± 1.3 $^{b, c, d}$	7.1 ± 0.5 ^{b}	11.7 ± 0.3 ^{c}	7.2 ± 0.5 $^{a, b}$	184.5 ± 13.0 ^{b}	228.4 ± 19.4 ^{b}
	1	22.6 ± 1.9 ^{d}	20.2 ± 1.4 $^{b, c}$	7.3 ± 0.5 ^{b}	9.1 ± 0.6 $^{c, d}$	7.6 ± 0.2 $^{a, b}$	196.2 ± 16.6 $^{a, b}$	240.4 ± 17.0 ^{b}
	4	43.0 ± 3.0 ^{d}	23.3 ± 2.0 $^{a, b}$	7.4 ± 0.5 ^{b}	47.3 ± 1.3 ^{b}	7.4 ± 0.5 $^{a, b}$	225.2 ± 15.9 ^{a}	310.6 ± 26.4 ^{a}
Pectinex®Ultra-Olio	0.5	247.3 ± 21.0 ^{c}	16.0 ± 1.1 $^{c, d}$	5.8 ± 0.4 ^{b}	53.5 ± 3.8 $^{a, b}$	8.4 ± 0.2 ^{a}	13.4 ± 1.1 ^{c}	97.1 ± 6.9 $^{c, d}$
	1	382.1 ± 27.0 ^{a}	22.1 ± 1.9 ^{b}	13.8 ± 1.0 ^{a}	55.6 ± 1.6 ^{a}	8.3 ± 0.6 ^{a}	15.2 ± 1.1 ^{c}	115.0 ± 9.8 ^{c}
	4	303.8 ± 25.8 ^{b}	29.0 ± 2.1 ^{a}	16.1 ± 1.1 ^{a}	55.2 ± 3.9 ^{a}	5.0 ± 0.1 ^{c}	23.2 ± 1.6 ^{c}	128.5 ± 9.1 ^{c}

^{a,b,c,d,e}Statistically significant differences between groups.
* **Total POS**: Σ POS-DP2, Di-GalA, POS-DP3, Tri-GalA and POS-DP > 3.

GalA (14.8 mg 100 mg^{-1} pectin) and neutral sugars (11.5 mg 100 mg^{-1} pectin), as expected, due to the high amount of added enzyme (6.75 U mL^{-1}), to attempt obtaining a large amount of low M_w POS. This elevated GalA release may also be due to their declared activity, pectin-lyase. In the other enzymatic preparations, although the enzyme dose was similar at 0.54, 0.63 and 0.88 U mL^{-1} of reaction mixture, the hydrolysis patterns of artichoke pectin were very different, probably because of their different main activities. Therefore, Cellulase from *A. niger* released a significantly major amount of GalA and Glucanex released mainly arabinose (in amounts significantly higher than Cellulase from *A. niger* and Pentopan but still lower than the ones obtained with Pectinex Olio). Lower hydrolysis rates were expected for Glucanex considering its main activity (glucanase) which may not produce significant ruptures in the pectin backbone. Results obtained with Cellulase from *A. niger* indicate significant pectinase secondary activity, even higher than its declared activity. On the contrary, the hydrolytic activity of Pentopan (α-1,4-endoxylanase) was very low showing a significantly lower monosaccharide release, probably due to a lower presence of xylose in artichoke pectin ramified chains. Interestingly, Cellulase from *A. niger* and Pectinex Olio also released 0.2-0.5 mg 100 mg^{-1} pectin of unidentified monosaccharides probably due to other secondary enzymatic activities. Neutral monosaccharides released from artichoke pectin hydrolysis, mainly arabinose and galactose, depict the relevance of a side chain of RG-I structures, possibly arabinans, galactans and arabinogalactans. As previously reported in Sabater, Corzo et al. (2018), artichoke pectin contains a higher percentage of neutral

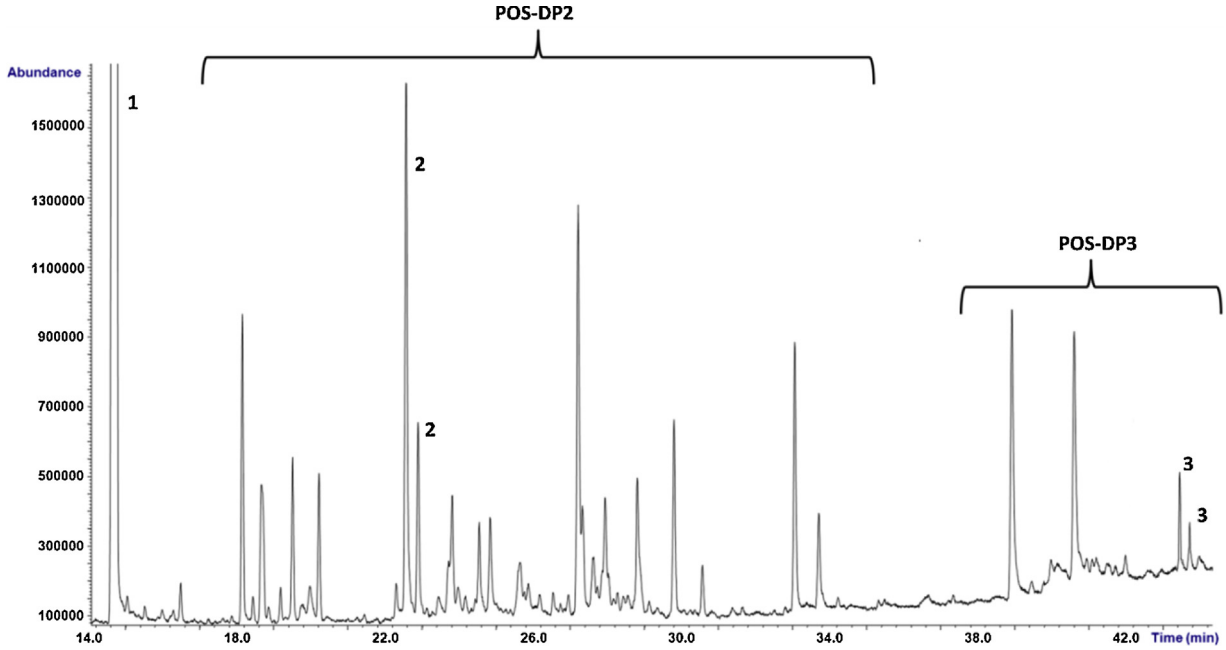


Fig. 5. GC–MS profile of hydrolysate of artichoke pectin obtained by incubation with cellulase from *A. niger*. Peaks: (1) Internal standard, (2) Digalacturonic acid (Di GalA), (3) Trigalacturonic acid (Tri-GalA). POS-DP2: unknown pectic disaccharides, POS-DP3: unknown pectic trisaccharides.

monosaccharides compared to pectin extracted from apple and citrus pectin (Bonnin, Garnier, & Ralet, 2014). Considering these differences, artichoke POS would be expected to show different structures to those derived from other pectin sources.

M_w distribution of enzymatic hydrolysates was determined by HPSEC-ELSD. Pectin from artichokes showed three main fragments with M_w of 660, 105 and 4.8 kDa, as has been previously reported (Sabater, Corzo et al., 2018) and after enzymatic treatments, differences in the chromatographic profiles were observed depending on the enzyme selected (Fig. 2). As can be seen, Pectinex Olio and Cellulase from *A. niger* (Fig. 2b) showed the most different patterns of POS average M_w distribution compared to the original product (Fig. 2a), producing the release of a wider variety of fragments and higher amounts of POS with M_w between 50 and 3 kDa (fractions with M_w < 3 kDa were not quantified due to coelution with other compounds present in enzymatic preparations). In contrast, chromatographic profiles of Pentopan hydrolysates showed hardly any modification with respect to artichoke pectin while Glucanex produced an important reduction de M_w producing modified pectins (M_w ~ 50 kDa) (Fig. 2a and c).

In general, average M_w distribution of POS produced with these preparations varied from 78 to 3.5 kDa (Fig. 3). Enzymatic hydrolysis with Cellulase from *A. niger* produced, after 1 h of reaction, POS of M_w around 14 kDa (60.3% of fragments determined) and the M_w decreased slightly after 4 h up to 11 kDa, while Pectinex Olio formed high amounts of POS with lower M_w (5.7–6.0 kDa, 77–80% of fragments determined). As has been stated, no changes were observed in the HPSEC-ELSD profiles of the other of hydrolysates (Fig. 2c). These results highlight the importance of enzymatic activity in POS structural characteristics and several studies reported the influence of enzyme and reaction conditions over POS formation (Babbar et al., 2016; Combo et al., 2012).

Differences in hydrolysis patterns were also observed in the HPAEC-PAD analysis. Fig. 4 shows a chromatographic profile of carbohydrates found in enzymatic hydrolysate of artichoke pectin after 4 h of reaction with enzyme preparations. Neutral sugars were eluted before acid sugars, GalA (peak 1), Di-GalA (peak 2), Tri-GalA (peak 3), and unknown acid POS with degree of polymerization (DP) of 2, 3 and > 3 were found. In comparison to retention times with commercial standards, these compounds could correspond to oligosaccharides with one/two molecules of GalA and one or more molecules of neutral sugars. Total acid POS formed with the four enzymes used were quantified (Table 2) and ranged from 33.4 to 310.6 mg g^{-1} pectin. Pectinex Olio released significantly higher amounts of GalA, Di-GalA and Tri-GalA. Interestingly, Cellulase from *A. niger* formed significant higher amounts of larger oligosaccharides. Maximum formation of POS was reached at 4 h of enzymatic hydrolysis, with a POS yield of 45.7, 63.7, 128.5 and 310.6 mg g^{-1} pectin for Pentopan, Glucanex, Pectinex Olio and Cellulase from *A. niger*, respectively.

3.2. GC–MS characterisation of POS-DP2 and POS-DP3 obtainment from artichoke pectin

To delve into structural characteristics of novel artichoke POS mixtures released during enzymatic hydrolysis, a GC–MS study of these oligosaccharides was carried out. In Fig. 5a CG-MS profile of hydrolysate of artichoke pectin using Cellulase from *A. niger* is shown. Di-GalA and Tri-GalA as well as unknown POS-DP2 and POS-DP 3 were detected. Similar GC–MS profiles have been obtained for the other enzymes used.

Interestingly, specific *m/z* ions 277, 321, 333 and 423, derived from β-cleavage fragmentation of uronic acids, as well as GalA characteristic ions *m/z* 332 and 540 were detected in Di-GalA MS spectra (Füzfai, Kovács, & Molnár-Perl, 2004; Petersson, 1974). Taking into account the scarce information found in the bibliography about the GC–MS spectra of this type of compound a mass spectral study employing three machine learning algorithms (multilayer perceptron, MLP; random forest, RF; and boosted logistic regression, BLR) was carried out. Therefore, full mass spectra of all disaccharides and trisaccharides and unknown POS found in the enzymatic hydrolysates were classified. The number of spectra included in this study is shown in Table 3. A total of 462 MS spectra were used for data analysis and classified according to the enzyme used, Glucanex-unknown POS (n = 104); Cellulase from *A. niger*-unknown POS (n = 116); Pentopan-unknown POS (n = 44); Pectinex Olio-unknown POS (n = 128) and different known pectic sugars (arabinose, rhamnose, xylose, galactose, GalA and Di- and Tri-GalA, n = 70).

For data analysis, 151 fragments in the range of *m/z* 61–546, which were statistically different among groups ($p < 0.05$) and might

Table 3
Number of GC–MS spectra (n = 462) of oligosaccharides obtained from GC–MS-EI analysis of hydrolysates from artichoke pectins included in machine learning study. GalA: galacturonic acid, Di-GalA: digalacturonic acid, Tri-GalA: trigalacturonic acid, POS: pectic-oligosaccharides.

Artichoke pectin + enzyme	Commercial standards	Number of hydrolysates	Number of GC-MS spectra					
			Known pectic monosaccharides	Di-GalA	Unknown POS-DP2	Tri-GalA	Unknown POS-DP3	Total
Glucanex® 200 G	–	4	–	8 (2 per hydrolysate)	60 (15 per hydrolysate)	8 (2 per hydrolysate)	44 (11 per hydrolysate)	16 known POS 104 unknown POS
Cellulase from *A. niger*	–	4	–	8 (2 per hydrolysate)	84 (21 per hydrolysate)	8 (2 per hydrolysate)	32 (8 per hydrolysate)	16 known POS 116 unknown POS
Pentopan® Mono-BG	–	4	–	8 (2 per hydrolysate)	24 (6 per hydrolysate)	8 (2 per hydrolysate)	20 (5 per hydrolysate)	16 known POS 44 unknown POS
Pectinex® Ultra-Olio	–	4	–	8 (2 per hydrolysate)	112 (28 per hydrolysate)	–	16 (4 per hydrolysate)	8 known POS 128 unknown POS
–	Standard	–	10[a]	2[b]	–	2[c]	–	14

[a] Known monosaccharides including galacturonic acid, arabinose, rhamnose, xylose and galactose.
[b] Digalacturonic acid standards.
[c] Trigalacturonic acid standards.

correspond to known POS ruptures, assessed by CFM-ID (Allen et al., 2016), were selected. Then, each MS spectra was decomposed and reconstructed using the DWT. Unknown POS were classified according to the enzyme used (Glucanex, Pentopan, Cellulase from *A. niger* and Pectinex Olio), and were also differentiated from known pectic sugars. These models were validated and tested on 30% of new samples. The training, 10-fold cross-validation and test rates were:

a) MLP: 100, 91.1 and 95.7%, respectively.
b) RF: 100, 97.8 and 100%, respectively.
c) BLR: 100, 98.1 and 100%, respectively.

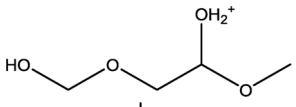
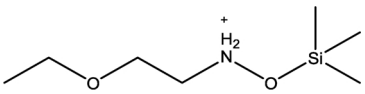
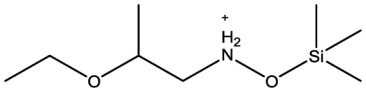
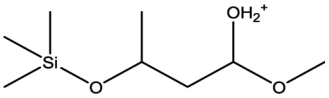
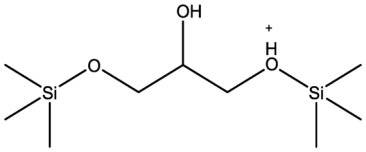
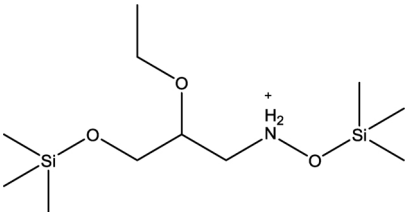
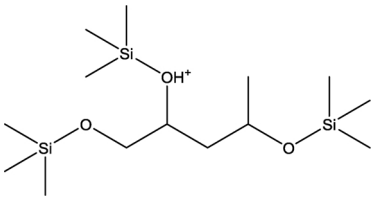
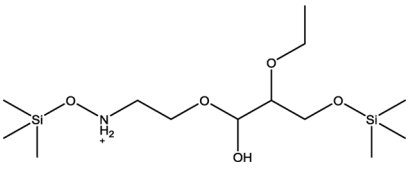
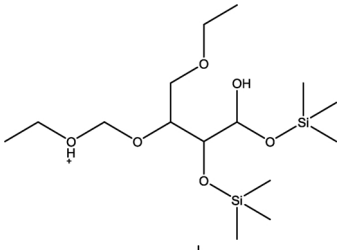
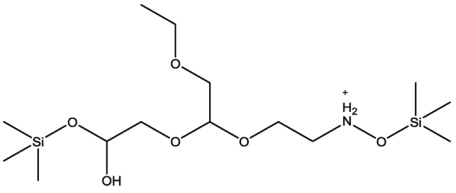
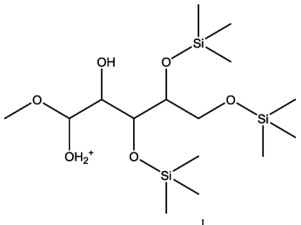
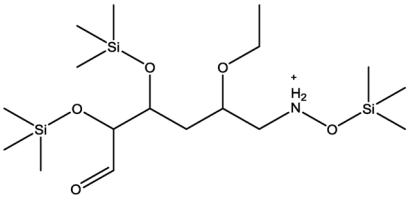
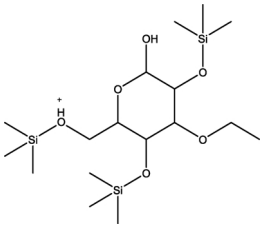
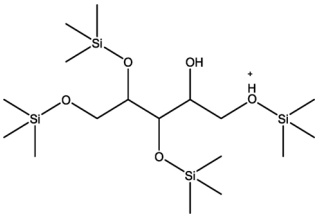
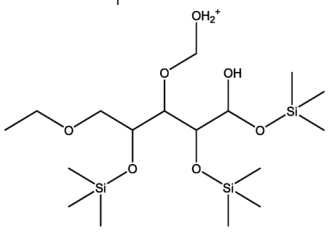
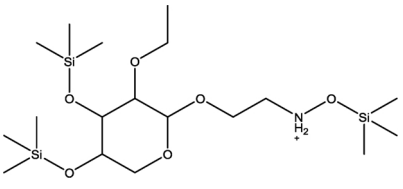
As can be observed, machine learning algorithms found a highly reproducible classification pattern that could be applied on new samples. DWT pre-processing allowed model prediction rates to increase up to 95–100%. When these models were tested on new samples, MLP showed false negative rates of 0.72, 1.45 and 2.17% for POS obtained with Glucanex, Pentopan and Pectinex Olio preparations, respectively. The false positive rates were 0.72% for known pectic sugars, also for POS obtained with cellulase from *A. niger*, Pentopan and Pectinex Olio were 0.72% and 1.60% for POS from Glucanex preparations. Although MLP showed high accuracy in its predictions (above 95%), RF and BLR correctly classified 100% of test samples. Therefore, RF and BLR showed 100% sensitivity, specificity and balanced accuracy values for each class. In contrast, MLP showed test sensitivity values between 85.7–96.7% for Pentopan, Pectinex Olio and Glucanex POS and 100% for Cellulase from *A. niger* POS and known sugars. MLP specificity values ranged from 98.2 to 99.2% for all the classes and balanced accuracy between sensitivity and specificity was 92.5, 95.7, 97.4, 99.5 and 99.6% for Pentopan, Pectinex Olio, Glucanex and Cellulase from *A. niger* POS and known sugars, respectively. Interestingly, MLP possessed higher specificity for POS classification, obtaining the lowest accuracy for Pentopan POS. However, all these values were above 90%. Higher accuracy for known sugars might be due to structural similarities among POS mixtures. The most influential *m/z* ions and their importance in each model were determined (Supplementary material Tables S1–S3).

In addition, overall accuracy and kappa values obtained with each model via their resampling distributions were compared (a comparative account is shown in Supplementary material Fig. S1). The accuracy indicates the number of instances that were classified correctly, while kappa is a more robust measure that takes into account the possibility of a correct classification by chance. Both values were similar in the three models. BLR and RF showed mean accuracy and kappa values of 0.97-0.98, significantly higher than those of MLP (0.91 and 0.89, respectively). In all three models, these metrics were high, indicating high predictive power.

To deepen into POS structural differences established by these models, a total 50 *m/z* values of the most relevant fragments were selected and probable chemical structures of these fragments were suggested employing CFM-ID code (Allen et al., 2016). Proposed structures of some interesting *m/z* ions (n = 20) are shown in Table 4, while the other structures of *m/z* ions (n = 30) are shown in Supplementary material Table S4.

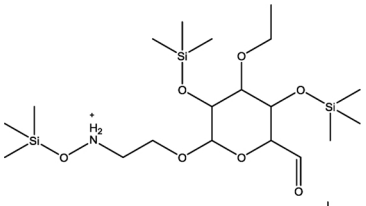
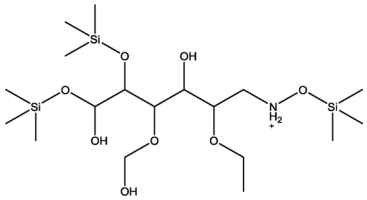
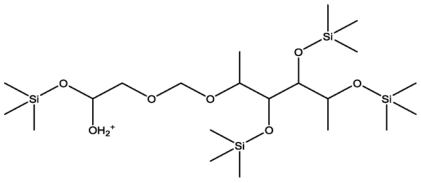
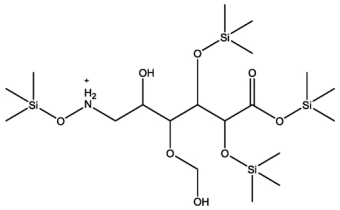
First, complete feasible fragments for candidate molecules (i.e. monosaccharides, POS-DP2 and DP3) were calculated by systematically breaking bonds within the molecule (Allen et al., 2015, 2016) in order to obtain a POS *in silico* fragmentation library. For monosaccharides, pectic neutral sugars (arabinose, rhamnose, xylose and galactose) and GalA (methylated/acetylated or not) were considered. The POS-DP2 included, taking into account more possible structures present in HG and RG-I, were Di-GalA (methylated or not), xylose-α(1,3)-GalA (methylated/acetylated or not), GalA-α(1,2)-rhamnose (acetylated or not), rhamnose-α(1,4)-GalA (acetylated or not), galactose-α(1,4)-rhamnose and arabinose-α(1,4)-rhamnose (Atmodjo et al., 2013; Mohnen, 2008). Tri-POS library was generated considering combinations of all POS-DP2

Table 4
Possible chemical structures of the most relevant *m/z* ions in POS classification given by machine learning algorithms, determined by CFM-ID. These ions correspond to TMSO fragments from pectic di- or trisaccharides.

m/z	Structure	*m/z*	Structure
123		178	
192		193	
237		280	
337		340	
369		370	
399		424	
425		441	
443		454	

(*continued on next page*)

Table 4 (*continued*)

m/z	Structure	*m/z*	Structure
482		488	
529		530	

structures and pectic monomers, following the main criteria reported by Atmodjo et al. (2013) and Mohnen (2008). Then, we proposed a chemical structure for each *m/z* ion by looking for these specific ruptures in our library for studying structural characteristics of known pectic sugars and unknown POS released during enzymatic hydrolysis. The abundances of several of the most relevant ions in machine learning classification were significantly higher ($p < 0.05$) in known pectic sugars released by the four enzymes, so fragments *m/z* 207 and 309 could be originated from all compounds studied. Ions *m/z* 311, 382, 414, 486, 498, 500, 501, 502, 513, 528, 532, 544, 545 and 546 were also higher in known pectic sugars and could also be derived from GalA-containing POS or GalA oligomers bonded by α(1,4) glycosidic linkages.

Then, statistically significant structural differences among unknown POS-DP2 and POS-DP3 released with the four preparations were studied:

Pentopan. Some *m/z* ions were significantly more abundant in unknown POS present in these hydrolysates: ion *m/z* 237 which could be derived from all pectic sugars and POS and *m/z* 193, 425 and 530 originated from xylose and POS ruptures. Other more specific ions higher in these POS were *m/z* 488 originated from POS containing rhamnose-α(1,4)-GalA (acetylated); ion *m/z* 399 is formed from GalA-α(1,2)-rhamnose, while *m/z* 529 is formed from POS containing rhamnose-α(1,4)-GalA (acetylated), GalA-α(1,2)-rhamnose (acetylated) or xylose-α(1,3)-GalA (methylated/acetylated or not); and ion *m/z* 454 which could also be formed from POS containing arabinose-α(1,4)-rhamnose.

Glucanex. Some fragments with higher abundance in unknown POS from Glucanex, *m/z* 280 derived from rhamnose-α(1,4)-GalA (acetylated), *m/z* 370 derived from GalA-α(1,2)-rhamnose (acetylated), and *m/z* 369 which could be derived from acetylated dimers of rhamnose-α(1,4)-GalA or GalA-α(1,2)-rhamnose as well as GalA-containing POS bonded by α(1,4) glycosidic linkages. Other important ions in these POS were *m/z* 178 and 192 which could also be originated from the rupture of xylose-α(1,3)-GalA (acetylated). Interestingly, *m/z* 123 derived from POS containing GalA (methylated) was significantly higher in Glucanex POS and lower in those obtained with cellulase from *A. niger* perhaps this enzyme produced slight demethylation.

Pectinex Olio. The MS study of unknown POS obtained with Pectinex Olio revealed that the abundances of *m/z* 337, which could be originated from all pectic sugars and all POS, and *m/z* 340, 424 and 443, specifically originated from rhamnose-α(1,4)-GalA, GalA-α(1,2)-rhamnose or xylose-α(1,3)-GalA (methylated/acetylated or not), were significantly higher in these hydrolysates.

Cellulase from A. niger. MS spectra of unknown POS formed with Cellulase from *A. niger* showed higher abundances of ions *m/z* 441, originated from galactose and all POS, and *m/z* 482, specifically derived from GalA-α(1,2)-rhamnose (acetylated) and xylose-α(1,3)-GalA (acetylated).

Other *m/z* ions which were relevant in more than one group were *m/z* 531 present in all POS containing GalA units and *m/z* 483 present in galactose-α(1,4)-rhamnose or GalA-α(1,2)-rhamnose. Similarly, abundances of *m/z* 353, 354, 457, 458, 471 and 472 were higher in unknown POS and could be derived from the rupture of most POS, with special importance for those containing rhamnose-α(1,4)-GalA, GalA-α(1,2)-rhamnose or xylose-α(1,3)-GalA. More specifically, *m/z* 225, 444, 469, 484 and 542, formed from the rupture of GalA and molecules containing GalA-α(1,4)-GalA dimers, were relevant in known pectic sugars and also in unknown POS from Pentopan and Glucanex preparations. Finally, *m/z* 439 formed from rhamnose-α(1,4)-GalA, GalA-α(1,2)-rhamnose and galactose-α(1,4)-rhamnose, was relevant in both hydrolysates with Pentopan or with Cellulase from *A. niger* preparations. At last, the dimer structures above referred may be also present in trisaccharides, but no trisaccharides specific *m/z* ions were relevant for POS classification.

The assignment of these glycosidic linkages was possible due to the most frequent structures found in pectins (Atmodjo et al., 2013; Mohnen, 2008).

As expected, *m/z* fragments possibly formed from the rupture of dimers and trimers containing neutral sugars and GalA (methylated/acetylated or not) units were relevant in unknown POS. In general terms, these results suggest that Glucanex and Pectinex Olio preparations produce POS that may contain rhamnose, xylose and GalA (methylated/acetylated or not). In addition, arabinose may also be present in POS structures produced by Pentopan. Moreover, Cellulase from *A. niger* may produce POS that contain neutral sugars such as galactose, rhamnose and xylose, and also acetylated GalA units.

A GC–MS spectral study might be considered as a first approach to determine structural characteristics of POS enzymatically obtained using commercial preparations with different activities. MS data modelling may lead to a better understanding of differences observed in the chromatographic profiles of these samples. It has been demonstrated that is possible to classify complex oligosaccharides according to the enzyme used for their obtainment, by extracting chemically relevant information from their full spectra. These models could be applied on new reaction mixtures containing novel oligosaccharides to evaluate the activity of different enzyme preparations considering their high prediction rates and allowing consistent structural information to be obtained. The ability to determine structural similarities and differences among novel POS can be of great importance to establish structure-function relationships which could be very useful to extrapolate results from biological assays.

4. Conclusions

Novel artichoke POS mixtures have been enzymatically obtained and characterised. Differences in the chromatographic profiles of POS were observed according to the enzyme used suggesting different structural properties, confirmed by several chromatographic techniques. Cellulase from *A. niger* and Pectinex® Ultra-Olio formed large amounts of POS, with M_w around 14 and 6.0 kDa, respectively while Glucanex produced high M_w POS and modified pectins, and Pentopan showed chromatographic profiles similar to those of original artichoke pectin. POS were analysed by HPAEC-PAD, showing that Cellulase from *A. niger* preparation produced the highest amount (310.6 mg g^{-1} pectin). Mass spectra of unknown POS-DP2 and -DP3 have been studied and classified using machine learning algorithms (multilayer perceptron, random forest and boosted logistic regression) obtaining high prediction rates on the test set. These models confirm structural differences observed in the hydrolysis profiles of commercial preparations with various enzymatic activities and could be used to establish structure-function relationships.

Acknowledgments

This work has been funded by MICINN of Spain, Projects AGL2014-53445-R and AGL2017-84614-C2-1-R and by the Spanish Danone Institute. Carlos Sabater thanks his FPU Predoc contract from Spanish MECD (FPU14/03619). Authors are also thankful to Riberebro (La Rioja, Spain) for kindly providing the artichoke by-products studied in this paper and Ramiro Martinez (Novozyme) for enzyme supply.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.carbpol.2018.10.054.

References

Aldrich, E. (2013). *Wavelets: A package of functions for computing wavelet filters, wavelet transforms and multiresolution analyses. R package version 0.3-0.* URL . . Last accessed: 30/07/18 https://CRAN.R-project.org/package=wavelets.

Allard, P. M., Péresse, T., Bisson, J., Gindro, K., Marcourt, L., Pham, V. C., et al. (2016). Integration of molecular networking and in-silico MS/MS fragmentation for natural products dereplication. *Analytical Chemistry, 88*, 3317–3323.

Allen, F., Greiner, R., & Wishart, D. (2015). Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification. *Metabolomics, 11*, 98–110.

Allen, F., Pon, A., Greiner, R., & Wishart, D. (2016). Computational prediction of electron ionization mass spectra to assist in GC/MS compound identification. *Analytical Chemistry, 88*, 7689–7697.

Atmodjo, M. A., Hao, Z. Y., & Mohnen, D. (2013). Evolving views of pectin biosynthesis. Ed. S.S. Merchant. *Annual Review of Plant Biology, 64*, 747–779.

Babbar, N., Dejonghe, W., Gatti, M., Sforza, S., & Elst, K. (2016). Pectic oligosaccharides from agricultural by-products: Production, characterization and health benefits. *Critical Reviews in Biotechnology, 36*, 594–606.

Babbar, N., Dejonghe, W., Sforza, S., & Elst, K. (2017). Enzymatic pectic oligosaccharides (POS) production from sugar beet pulp using response surface methodology. *Journal of Food Science and Technology, 54*, 3707–3715.

Bergmeir, C., & Benitez, J. M. (2012). Neural networks in R using the Stuttgart neural network simulator: RSNNS. *Journal of Statistical Software, 46*, 1–26. URL http://www.jstatsoft.org/v46/i07/ . Last accessed: 30/07/18 .

Boccard, J., & Rudaz, S. (2014). Harnessing the complexity of metabolomic data with chemometrics. *Journal of Chemometrics, 28*, 1–9.

Bonnin, E., Garnier, C., & Ralet, M. C. (2014). Pectin-modifying enzymes and pectin derived materials: Applications and impacts. *Applied Microbiology and Biotechnology, 98*(2), 519–532.

Combo, A. M. M., Aguedo, M., Goffin, D., Wathelet, B., & Paquot, M. (2012). Enzymatic production of pectic oligosaccharides from polygalacturonic acid with commercial pectinase preparations. *Food and Bioproducts Processing, 90*, 588–596.

Combo, A. M. M., Aguedo, M., Quiévy, N., Danthine, S., Goffin, D., Jacquet, N., et al. (2013). Characterization of sugar beet pectic-derived oligosaccharides obtained by enzymatic hydrolysis. *International Journal of Biological Macromolecules, 52*, 148–156.

da Moura, F. A., Macagnan, F. T., & da Silva, L. P. (2015). Oligosaccharide production by hydrolysis of polysaccharides: A review. *International Journal of Food Science & Technology, 50*, 275–281.

Füzfai, Z., Kovács, E., & Molnár-Perl, I. (2004). Identification and quantitation of the main constituents of sour cherries: Simultaneously, as their trimethylsilyl derivatives, by gas chromatography-mass spectrometry. *Chromatographia, 60*, S143–S151.

Gertheiss, J., & Tutz, G. (2009). Supervised feature selection in mass spectrometry-based proteomic profiling by blockwise boosting. *Bioinformatics, 25*, 1076–1077.

Geurts, P., Fillet, M., De Seny, D., Meuwis, M. A., Malaise, M., Merville, M. P., et al. (2005). Proteomic mass spectra classification using decision tree based ensemble methods. *Bioinformatics, 21*, 3138–3145.

Gómez, B., Gullón, B., Yáñez, R., Parajó, J. C., & Alonso, J. L. (2013). Pectic oligosacharides from lemon peel wastes: Production, purification, and chemical characterization. *Journal of Agricultural and Food Chemistry, 61*(42), 10043–10053.

Gómez, B., Yañez, R., Parajo, J. C., & Alonso, J. L. (2016). Production of pectin-derived oligosaccharides from lemon peels by extraction, enzymatic hydrolysis and membrane filtration. *Journal of Chemical Technology & Biotechnology, 91*, 234–247.

Gosav, S., Praisler, M., & Birsa, M. L. (2011). Principal component analysis coupled with artificial neural networks—A combined technique classifying small molecular structures using a concatenated spectral database. *International Journal of Molecular Sciences, 12*, 6668–6684.

Gullón, B., Gómez, B., Martínez-Sabajanes, M., Yáñez, R., Parajó, J. C., & Alonso, J. L. (2013). Pectic oligosaccharides: Manufacture and functional properties. *Trends in Food Science & Technology, 30*, 153–161.

Herbstreith, & Fox (2018). *The specialists for pectin.* URL . . Last accessed: 30/07/18 http://www.herbstreith-fox.de/index.php?id=54&L=1.

Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., & MacCoss, M. J. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature Methods, 4*, 923–925.

Leijdekkers, A. G., Huang, J. H., Bakx, E. J., Gruppen, H., & Schols, H. A. (2015). Identification of novel isomeric pectic oligosaccharides using hydrophilic interaction chromatography coupled to traveling-wave ion mobility mass spectrometry. *Carbohydrate Research, 404*, 1–8.

Li, X., Li, J., & Yao, X. (2007). A wavelet-based data pre-processing analysis approach in mass spectrometry. *Computers in Biology and Medicine, 37*, 509–516.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2*, 18–22.

Lim, D. K., Long, N. P., Mo, C., Dong, Z., Cui, L., Kim, G., et al. (2017). Combination of mass spectrometry-based targeted lipidomics and supervised machine learning algorithms in detecting adulterated admixtures of white rice. *Food Research International, 100*, 814–821.

Lin, Z., Gonçalves, C. M. V., Dai, L., Lu, H. M., Huang, J. H., Ji, H., et al. (2014). Exploring metabolic syndrome serum profiling based on gas chromatography mass spectrometry and random forest models. *Analytica Chimica Acta, 827*, 22–27.

Maric, M., Grassino, A. N., Zhu, Z. Z., Barba, F. J., Brncic, M., & Brncic, S. R. (2018). An overview of the traditional and innovative approaches for pectin extraction from plant food wastes and by-products: Ultrasound-, microwaves-, and enzyme-assisted extraction. *Trends in Food Science & Technology, 76*, 28–37.

Martínez, M., Gullón, B., Yanez, R., Alonso, J. L., & Parajó, J. C. (2009). Direct enzymatic production of oligosaccharide mixtures from sugar beet pulp: Experimental evaluation and mathematical modeling. *Journal of Agricultural and Food Chemistry, 57*, 5510–5517.

Mohnen, D. (2008). Pectin structure and biosynthesis. *Current Opinion in Plant Biology, 11*, 266–277.

Ognyanov, M., Remoroza, C., Schols, H. A., Georgiev, Y., Kratchanova, M., & Kratchanov, C. (2016). Isolation and structure elucidation of pectic polysaccharide from rose hip fruits (*Rosa canina* L.). *Carbohydrate Polymers, 151*, 803–811.

Petersson, G. (1974). Gas-chromatographic analysis of sugars and related hydroxy acids as acyclic oxime and ester trimethylsilyl derivatives. *Carbohydrate Research, 33*, 47–61.

Rossel, S., & Martínez Arbizu, P. (2018). Automatic specimen identification of Harpacticoids (Crustacea: Copepoda) using Random Forest and MALDI-TOF mass spectra, including a post hoc test for false positive discovery. *Methods in Ecology and Evolution, 9*, 1421–1434.

Ruttkies, C., Schymanski, E. L., Wolf, S., Hollender, J., & Neumann, S. (2016). MetFrag relaunched: Incorporating strategies beyond *in silico* fragmentation. *Journal of Cheminformatics, 8*, 3.

Sabater, C., Corzo, N., Olano, A., & Montilla, A. (2018). Enzymatic extraction of pectin from artichoke (*Cynara scolymus L.*) byproducts using Celluclast®1.5L. *Carbohydrate Polymers, 190*, 43–49.

Sabater, C., Montilla, A., Ovejero, A., Prodanov, M., Olano, A., & Corzo, N. (2018). Furosine and HMF determination in prebiotic-supplemented infant formula from Spanish market. *Journal of Food Composition and Analysis, 66*, 65–73.

Schymanski, E. L., Ruttkies, C., Krauss, M., Brouard, C., Kind, T., Dührkop, K., et al. (2017). Critical assessment of small molecule identification 2016: Automated methods. *Journal of Cheminformatics, 9*, 1–21.

Tuszynski, J. (2014). *caTools: Tools: Moving window statistics, GIF, Base64, ROC AUC, etc.. R package version 1.17.1.* Last accessed: 30/07/18 https://CRAN.R-project.org/package=caTools.

Uarrota, V. G., Moresco, R., Coelho, B., da Costa Nunes, E., Peruch, L. A. M., de Oliveira Neubert, E., et al. (2014). Metabolomics combined with chemometric tools (PCA, HCA, PLS-DA and SVM) for screening cassava (Manihot esculenta Crantz) roots during postharvest physiological deterioration. *Food Chemistry, 161*, 67–78.

Varmuza, K., He, P., & Fang, K. T. (2003). Boosting applied to classification of mass spectral data. *Journal of Data Science, 1*, 391–404.

Xia, J. M., Wu, X. J., & Yuan, Y. J. (2007). Integration of wavelet transform with PCA and ANN for metabolomics data-mining. *Metabolomics, 3*, 531–537.

Yasui, Y., Pepe, M., Thompson, M. L., Adam, B. L., Wright, G. L., Jr, Qu, Y., et al. (2003). A data-analytic strategy for protein biomarker discovery: Profiling of high-dimensional proteomic data for cancer detection. *Biostatistics, 4*, 449–463.

Yi, L., Dong, N., Yun, Y., Deng, B., Ren, D., Liu, S., et al. (2016). Chemometric methods in data processing of mass spectrometry-based metabolomics: A review. *Analytica Chimica Acta, 914*, 17–34.

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

Cite This: J. Agric. Food Chem. 2019, 67, 7435−7447

pubs.acs.org/JAFC

Enzymatic Production and Characterization of Pectic Oligosaccharides Derived from Citrus and Apple Pectins: A GC-MS Study Using Random Forests and Association Rule Learning

Carlos Sabater, Alvaro Ferreira-Lazarte, Antonia Montilla,* and Nieves Corzo

Institute of Food Science Research, CIAL (CSIC-UAM), CEI (UAM+CSIC), C/Nicolás Cabrera 9, Madrid 28049, Spain

Supporting Information

ABSTRACT: Pectic oligosaccharides (POS) from citrus and apple pectin hydrolysis using ViscozymeL and Glucanex200G have been obtained. According to the results, maximum POS formation was achieved from citrus pectin after 30 min of hydrolysis with ViscozymeL, with a yield of 652 mg g^{-1} and average molecular mass (M_w) of 0.8−2.5 kDa, while with Glucanex200G, the yield was 518 mg g^{-1} and M_w was 0.8−7.1 kDa. Digalacturonic and trigalacturonic acids were identified among other low M_w compounds as di- and tri-POS. In addition, differences in GC-MS spectra of all oligosaccharides found in the hydrolysates were studied by employing random forests and other algorithms to identify structural differences between the obtained POS, and high prediction rates were shown for new samples. Chemical structures were proposed for some influential m/z ions, and 12 association rules that explain differences according to pectin and enzyme origin were built. This information could be used to establish structure−function relationships of POS.

KEYWORDS: *pectic oligosaccharides, ViscozymeL, Glucanex200G, machine learning, in silico fragmentation*

INTRODUCTION

Currently, there is an overall interest in obtaining novel oligosaccharides with improved functional properties using plant cell wall polysaccharides as raw material.[1] Pectic oligosaccharides (POS) have been proposed as a new class of prebiotics due to their health-promoting activities including prevention of the adhesion of uropathogenic microorganisms, stimulation of apoptosis of human colonic adenocarcinoma cells, or cardiovascular protection.[2]

POS can be obtained from depolymerization of raw materials or purified pectins mainly extracted from apple, citrus peel, and sugar beet byproducts, using enzymes, chemical reagents (acids), or combined methods.[3,4] Also, other substrates, such as onion skins,[5] rapeseed cake,[6] and artichoke pectin,[7] have been used to obtain POS.

On the other hand, pectins are one of the most structurally complex families of polysaccharides in nature containing three major pectic polysaccharides: homogalacturonan (HG) and rhamnogalacturonan I and II (RG-I and RG-II). HG, the most abundant domain, is a homopolymer of D-galacturonic acid (GalA) with α-1,4 linkages (GalA ∼ 65% of pectin). This polymer is partially methyl-esterified at the *O*-6 position and may be *O*-acetylated at *O*-2 or *O*-3; also, some GalA units may be substituted at *O*-3 with β-D-xylose, forming xylogalacturonan (XGA). The second most important domain is RG-I, with alternate sequences of GalA and L-rhamnosyl (Rha) residues in its backbone [-4)-α-D-Gal*p*A-(1,2)-α-L-Rha*p*-(1-]$_n$. In this domain, rhamnose may be substituted at *O*-4 with linear or branched oligosaccharides of arabinose (arabinans), galactose (galactans), or both (arabinogalactans). The extent of branching depends on neutral sugar content, so it may vary according to pectin source. RG-II, consisting of up to 12 different types of sugars in over 20 different linkages, constitutes another minor domain of pectin.[8] Due to complexity of pectin structure, the resulting POS can have varied chemical structure, which will influence their bioactive properties.

To date, most studied food-grade POS have been produced by commercial enzymes, mostly pectinases, usually produced by *Aspergillus* spp.[9,10] These POS can present very different chemical structures depending on the type of enzyme, reaction conditions, and substrate used.[1]

Different analytical techniques are employed to identify and characterize POS: high performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD),[3,7] high performance size-exclusion liquid chromatography with refractive index detector (HPSEC-RID),[11] high performance size-exclusion liquid chromatography with evaporative light scattering detector (HPSEC-ELSD),[7] and hydrophilic interaction liquid chromatography-mass spectrometry (HILIC-MS),[12] the latter being one of the most used. However, GC-MS can be of great interest because it generates massive amounts of high-dimensional data.[7] To interpret GC-MS-EI results, *in silico* fragmentation methods have been developed.[13,14]

Another important step in MS data analysis involves data modeling to find valuable information about chemical structures. With this aim, several machine learning algorithms have been applied.[15] Some common algorithms for classifying metabolomics and food science data include random forests (RF) and artificial neural networks (ANN).[16−18] These models

Received: February 7, 2019
Revised: May 29, 2019
Accepted: June 4, 2019
Published: June 4, 2019

DOI: 10.1021/acs.jafc.9b00930
J. Agric. Food Chem. 2019, 67, 7435−7447

could be of great importance for sample characterization. ANN have been applied on GC-MS data to classify small molecular structures[19] and compared to RF and boosted logistic regression (BLR) to characterize POS substructures obtained from low methyl-esterified artichoke pectin.[7] However, more studies are needed for determining GC-MS structural differences of POS, for example, POS obtained from an HG type structure and highly methyl-esterified pectin. In addition, in a few studies, association rules between m/z ions for establishing fragmentation patterns that could explain these differences have been calculated.

Therefore, the aim of this work was to characterize novel POS obtained from the hydrolysis of polygalacturonic acid (PGA) and highly methyl-esterified pectin (from citrus and apple) using two commercial enzymes with different activities, endopolygalacturonase (endo-PG) and others, including arabinosidase and galactosidase, through their GC-MS spectra. For this purpose, spectral data modeling has been improved to gain a better understanding of oligosaccharide fragment relationships. Machine learning algorithms (RF, BLR, and ANN) were computed, and most relevant fragments were selected to build association rules that explain some fragmentation patterns and structural differences of POS depending on the enzyme and sample types.

MATERIALS AND METHODS

Standards and Reagents. Analytical reference substances such as sucrose, L-arabinose, D-xylose, L-rhamnose, D-galactose, D-mannose, D-glucose, D-fructose, sucrose, kestose, nystose, galacturonic acid (GalA), digalacturonic acid (Di-GalA), PGA (purity >85%, GalA > 96%, degree of methyl-esterification (DM) of 82.0 ± 0.0%, determined by FT-IR), and phenyl-β-glucoside were purchased from Sigma-Aldrich (Steinheim, Germany). Trigalacturonic acid (Tri-GalA) was from Carbosynth (Compton, UK). Commercial apple pectin with a GalA > 74%, according to the label, and DM of 83.0 ± 0.5% was acquired from Sigma-Aldrich. Commercial citrus pectin with a GalA > 74% and a molecular mass (M_w) ranging from 22 to 400 kDa, according to the label, and DM of 71.0 ± 1.4% was acquired from Acofarma (Terrassa, Barcelona). Commercial enzyme preparations Glucanex200G (Glucanex, β-glucanase from *Trichoderma harzianum*) and ViscozymeL (Viscozyme, endo-1,3(4)-β-glucanase from *A. aculeatus*) were a generous gift from Novozymes.

Enzyme Characterization. Enzyme characterization assays were carried out using the method of DNS with a 1% (w/v) PGA solution dissolved in 0.05 M acetate buffer (pH 4.5) as substrate and 1 mg mL^{-1} of Glucanex or 1 μL mL^{-1} of Viscozyme. Samples were incubated at 50 °C for 2 min. One unit of pectinase activity was defined as the amount of enzyme required to release 1 μmol of GalA per min at 50 °C. Pectinase activity was 3778 U mL^{-1} and 739 U g^{-1} for Viscozyme and Glucanex preparations, respectively.

Total protein was determined by the Bradford method as 203 mg mL^{-1} and 172 mg g^{-1} for Viscozyme and Glucanex, respectively. Specific pectinase activity was 18.6 and 4.3 U mg^{-1} protein for Viscozyme and Glucanex preparations, respectively.

Pectin Hydrolysis. Two pectin depolymerization experiments were carried out (Figure S1).

Hydrolysis of Pectins To Know Monosaccharide Composition. Citrus and apple pectins were submitted to a chemical and enzymatic hydrolysis.

For the chemical hydrolysis, 10 mg of pectin was hydrolyzed with 500 μL of 2 M trifluoroacetic acid (TFA) at 110 °C under inert conditions for 4 h.[20]

For the enzymatic hydrolysis, 1% (w/v) commercial pectin solutions in 0.05 M acetate buffer (pH 4.5) were hydrolyzed with 90 U mL^{-1} of Viscozyme. Samples were incubated in a thermomixer at 50 °C and 750 rpm for 24 h.[20]

Enzymatic Obtainment of Pectic Oligosaccharides (POS). POS were obtained by enzymatic hydrolysis of 1% (w/v) commercial pectin solutions and PGA, prepared in 0.05 M acetate buffer (pH 4.5), using 16 and 4 U mL^{-1} of Viscozyme and 16 U mL^{-1} of Glucanex incubated in an orbital shaker at 50 °C and 750 rpm. Aliquots (1 mL) were withdrawn from the reaction mixture at short (30 min) and long times (24 h) and immediately immersed in boiling water for 5 min to inactivate the enzyme. Samples were stored at −18 °C for subsequent analysis. Enzymatic reactions were carried out in duplicate, and analyses were repeated at least twice for each enzymatic treatment.

Chromatographic Determination of Carbohydrates. *Gas Chromatography with Flame Ionization Detector (GC-FID) and Mass Spectrometry (GC-MS).* Carbohydrates were analyzed as trimethylsilylated oximes (TMSO). GC-FID and GC-MS analyses were performed according to the method described by Sabater et al.[7]

High Performance Size-Exclusion Liquid Chromatography with Refractive Index Detector (HPSEC-RID). The M_w distribution of carbohydrates in the reaction mixtures was determined by HPSEC-RID according to Muñoz-Almagro et al.[11]

High Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD). POS formed after 30 min of enzymatic hydrolysis were analyzed by HPAEC-PAD following the method of Sabater et al.[7]

Statistical Analyses. ANOVA tests and Tukey's test for $p < 0.05$ were applied to all data generated.

Study of GC-MS Spectra. In order to study the structural differences between POS (di- and trisaccharides) obtained with enzymatic preparations, a comprehensive GC-MS study was carried out (Figure 1). Full spectra (n = 212) of unknown POS and known pectic sugars (arabinose, rhamnose, xylose, galactose, GalA, Di-GalA, and Tri-GalA) present in reaction mixtures (Table S1) were analyzed and classified according to three different criteria: (i) pectin type, (ii) enzyme type, and (iii) pectin and enzyme types. For this purpose, three machine learning models were used and compared to reinforce the results from each individual classification: random forests (RF), boosted logistic regression (BLR), and artificial neural networks (multilayer perceptron, MLP).

Signal Processing. Discrete wavelet transform (DWT) was applied to reduce spectral noise and, subsequently, increase model performance according to previous studies.[7] A 14 level DWT was computed. In the threshold step, DWT coefficients under the 15% percentile were removed to denoise the signal.

Classification Studies. Known pectic sugars and unknown POS preprocessed GC-MS spectra were introduced as inputs in three classification studies. RF and BLR were selected for the first two (i, ii) while MLP was also included in the third one (iii) to compare the predictive power. The following hyperparameters were tuned to compute the models:

RF: A RF model was built with 500 trees and 2 variables tried at each split for the first study while two RF with 1000 trees and 2 variables tried at each split were computed for the rest.

BLR: A BLR consisting of 21 iterations was computed for all studies.

MLP: A MLP was built with 1 hidden layer consisting of 41 hidden neurons. The activation function was logistic.

All the models were trained with 70% of the data, 10-fold cross-validated, and then tested with 30% of the data from each class (corresponding to new samples). Predictive power of machine learning algorithms was also compared to principal components analysis (PCA) and partial least-squares-discriminant analysis (PLS-DA). Variable importance analysis was carried out to determine the most influential m/z ions in each model according to Sabater et al.[7]

Influential m/z Selection. Most relevant and nonredundant m/z ions in each classification were selected by employing RF models for backward variable elimination. The least important variables were successively eliminated using the out of bag (OOB) error as the minimization criterion.

In Silico Fragmentation. After determining the most important m/z ions in POS classification, chemical structures of some fragments

DOI: 10.1021/acs.jafc.9b00930
J. Agric. Food Chem. 2019, 67, 7435−7447

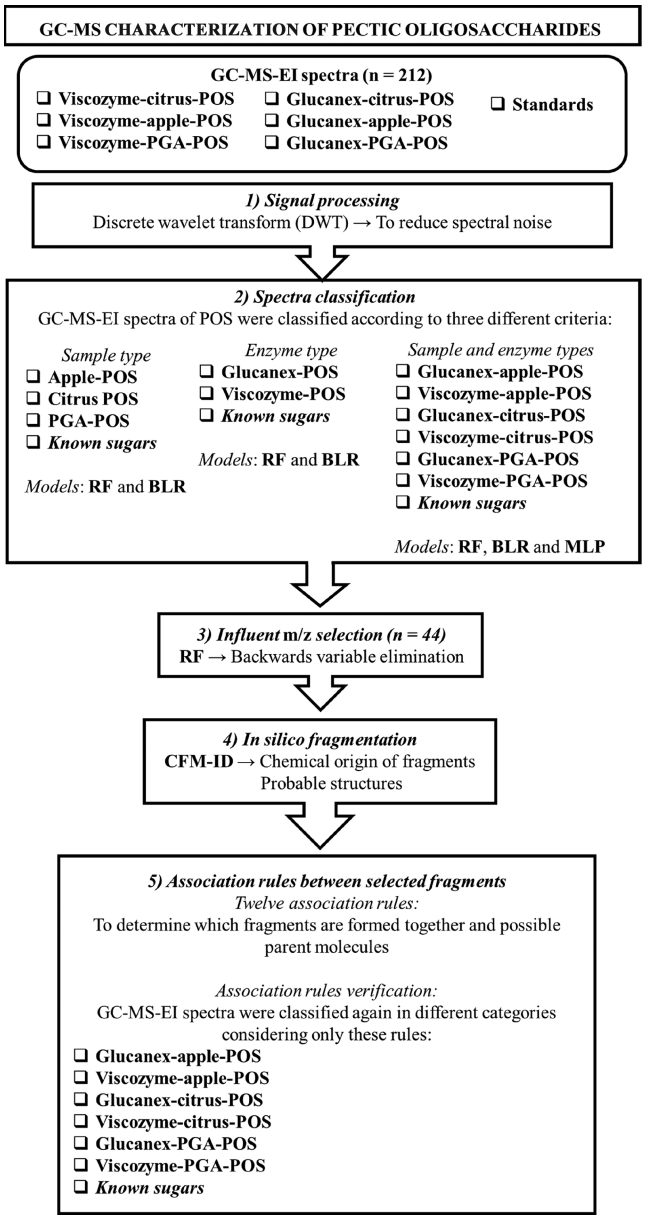


Figure 1. Schematic representation of the GC-MS-EI computational study.

have been proposed by employing the competitive fragmentation modeling source code (CFM-ID) developed by Allen et al.[14] GC-MS-EI *in silico* fragmentation patterns of POS structures, described in the bibliography,[8,21] were determined and compared to those of enzymatically obtained POS.[7]

Association Rules between Fragments. Association rule-based classification of GC-MS spectra considering relevant fragments selected by RF was carried out. Twelve association rules that could explain differences in the fragmentation pattern of POS from different enzymatic and pectin origin were established. This model was trained with 70% of the data and tested with 30% of new samples.

All statistical analyses were computed on R v3.5.0. DWT was performed using the wavelets package.[22] MLP was computed using the RSNNS package,[23] and RF classification was performed with the Random Forest package.[24] For BLR, the caTools package was employed.[25] Association rules were computed using the arulesCBA package.[26]

RESULTS AND DISCUSSION

Monomeric Characterization of Pectin. The monosaccharide composition of commercial citrus and apple pectins was determined by GC-FID in acid and enzymatic hydrolysates; Figure S2 shows a chromatogram of the carbohydrates found in xylose, arabinose, rhamnose, galactose, mannose, glucose, and GalA. The quantification of pectic monosaccharides is shown in Table 1. The results showed that chemical hydrolysis of pectin at high temperatures may be chosen in order to quantify neutral sugars, as it has been suggested by Soria et al.,[27] because the enzymatic hydrolysis did not lead to a complete release of arabinose and rhamnose. Gama et al.[28] determined a low α-L-arabinofuranosidase activity in Viscozyme during enzymatic hydrolysis of apple pomace. However, the enzymatic hydrolysis was most appropriate for quantifying GalA due to its thermolabile nature. Thus, enzymatic hydrolysis of pectin using the Viscozyme preparation at long times (24 h) and acid hydrolysis (4 h) allowed one to determine the GalA and neutral sugars, respectively, providing a better characterization of pectin.[20] Other authors also have determined the GalA and neutral sugars in enzymatic and acid hydrolysates, respectively.[29]

Taking into account the data shown in Table 1, the GalA content was the most abundant in both pectins, corresponding to 89.6% and 89.3% for apple and citrus pectin, respectively. Considering acid hydrolysis data, the former contained more neutral monosaccharides in its structure, i.e., arabinose and rhamnose, depicting the relevance of rhamnogalacturonan (RG) structures. However, xylose contents obtained with these two treatments were similar. The presence of high amounts of arabinose and xylose was also reported in apple pectin by the larger presence of xylo-galacturonan and arabino-galactans.[8] Glucose and mannose are not pectic carbohydrates, and they were present due to a possible extraction from cellulose and hemicelluloses including mannans during pectin obtainment.

Table 1. Monosaccharide Content (mg g^{-1} of Pectin) after Acid (TFA 2 M, 110 °C, 4 h) and Enzymatic (Viscozyme 90 U mL^{-1}, pH 4.5, 50 °C, 24 h) Hydrolysis of Citrus and Apple Pectins[a]

	citrus pectin		apple pectin	
carbohydrates (mg g^{-1} pectin)	acid hydrolysis	enzymatic hydrolysis	acid hydrolysis	enzymatic hydrolysis
xylose	0.9 ± 0.1 b	1.0 ± 0.3 b	5.6 ± 2.1 a	3.4 ± 2.1 a,b
arabinose	27.1 ± 0.9 b	7.3 ± 0.6 c	65.7 ± 6.8 a	21.9 ± 0.6 b
rhamnose (Rha)	21.1 ± 0.1 a	14.8 ± 0.4 b	23.9 ± 2.4 a	13.5 ± 0.4 b
galactose	69.5 ± 5.5 a	82.0 ± 3.5 a	53.1 ± 2.7 b	43.4 ± 1.7 b
galacturonic acid (GalA)	656.1 ± 35.6 a,b	875.0 ± 24.2 a	631.3 ± 97.5 b	704.5 ± 31.6 a,b
GalA/Rha	31 c	59 a	26 d	52 b

[a]Letters indicate statistically significant differences between groups.

DOI: 10.1021/acs.jafc.9b00930
J. Agric. Food Chem. 2019, 67, 7435−7447

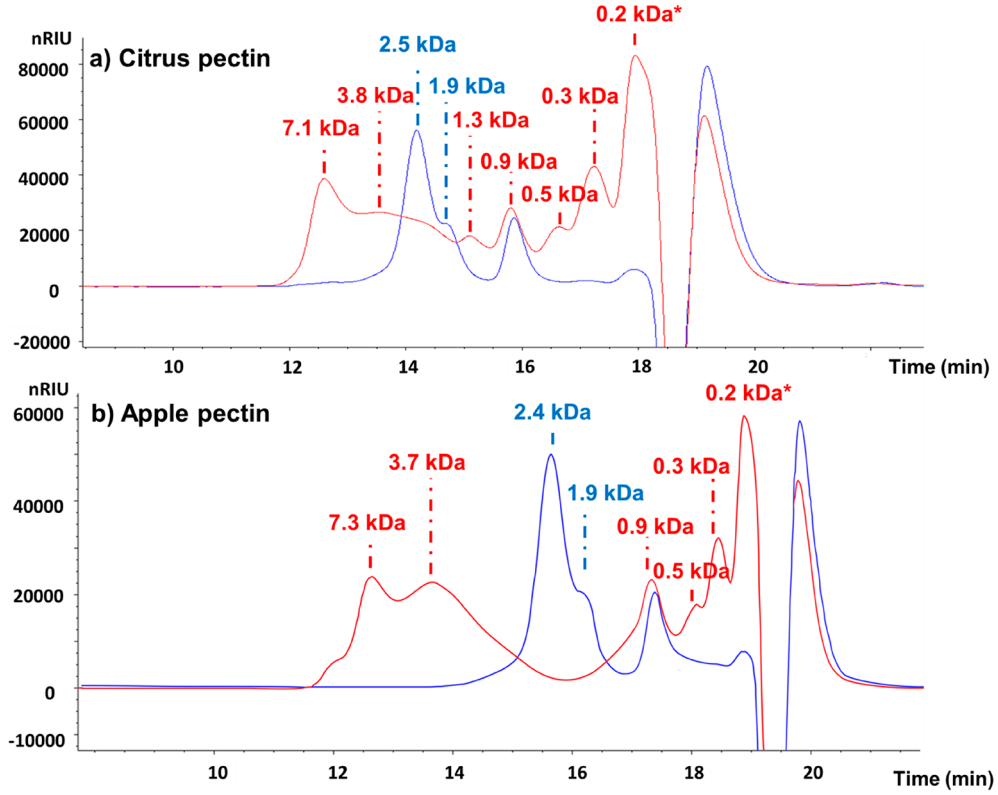


Figure 2. HPSEC-RID profiles of hydrolyzed citrus (a) and apple (b) pectins (30 min at 50 °C) using 16 U mL^{-1} of Glucanex (red) and 4 U mL^{-1} of Viscozyme (blue) preparations. *Compound (0.2 kDa) present in the Glucanex preparation.

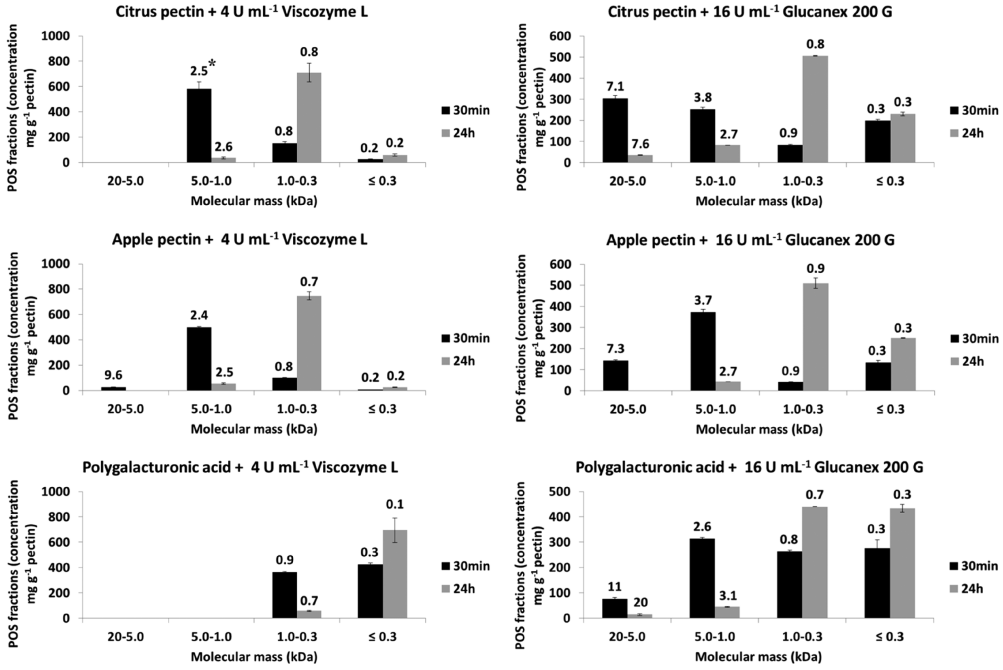


Figure 3. Molecular mass distribution of POS (HPSEC-RID) obtained from hydrolysates of commercial citrus and apple pectins and polygalacturonic acid (as pectin standard) using Glucanex (16 U mL^{-1}) and Viscozyme (4 U mL^{-1}) for 30 min and 24 h. *Average molecular mass of each region.

Production of Pectic Oligosaccharides (POS) Derived from Citrus and Apple Pectins. Considering that enzymatic hydrolysis is one of the best methodologies for obtaining POS, in this study, two commercial enzyme preparations were compared, Viscozyme and Glucanex with high and low polygalacturonase activity, respectively. Depolymerization of citrus and apple pectins and PGA (used as reference substrate) was carried out using Viscozyme (4 and 16 U mL^{-1}) and Glucanex (16 U mL^{-1}). Production of POS was followed by HPSEC-RID. When enzymatic hydrolysis of citrus and apple

DOI: 10.1021/acs.jafc.9b00930
J. Agric. Food Chem. 2019, 67, 7435−7447

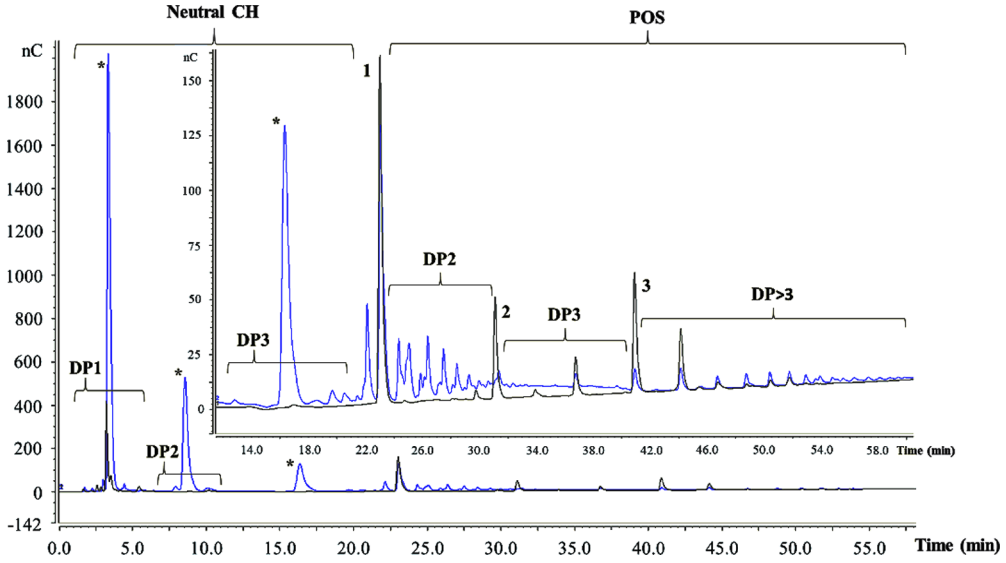


Figure 4. HPAEC-PAD profiles of carbohydrates found in hydrolyzed citrus pectin with 16 U mL^{-1} Glucanex200G (blue) and 4 U mL^{-1} Viscozyme (black) preparations for 30 min. DP: degree of polymerization; CH: carbohydrates; POS: pectic oligosaccharides. Peaks: (1) galacturonic acid, (2) digalacturonic acid, and (3) trigalacturonic acid. (*) Neutral carbohydrates present in the Glucanex200G preparation were not quantified.

Table 2. HPAEC-PAD Quantification (mg g^{-1} Pectin) of Pectic Oligosaccharides (POS) in Hydrolysates (Reaction Time: 30 min) of Studied Pectins and PGA (Polygalacturonic Acid)[a]

		POS (mg g^{-1} pectin)						
enzymes	pectin	GalA (peak4)	disaccharides with GalA	Di-GalA (peak5)	trisaccharides with GalA	Tri-GalA (peak6)	tetrasacchrides (or >3) with GalA	total POS[b]
ViscozymeL (4 U mL^{-1})	citrus	123 ± 5 c		166 ± 12 c	21 ± 1 b	67 ± 5 c	399 ± 39 a	652 ± 46 a,b
	apple	92 ± 7 c,d		125 ± 4 c	19 ± 1 b	52 ± 4 c,d	321 ± 23 a	517 ± 51 b,c
	PGA	200 ± 8 b		352 ± 25 a	21 ± 1 b	338 ± 24 a	56 ± 6 d	767 ± 54 a
Glucanex200G (16 U mL^{-1})	citrus	93 ± 7 c,d	238 ± 10 a	29 ± 1 d	16 ± 1 b	18 ± 1 d	217 ± 15 b	518 ± 51 b,c
	apple	60 ± 3 d	152 ± 11 b	19 ± 1 d	13 ± 0 b	11 ± 1 d	139 ± 14 b,c	335 ± 24 c
	PGA	319 ± 23 a	18 ± 1 c	247 ± 7 b	75 ± 5 a	190 ± 13 b	111 ± 8 c,d	641 ± 63 a,b

[a]GalA: galacturonic acid; Di-GalA: digalacturonic acid; Tri-GalA: trigalacturonic acid; POS: pectic oligosaccharides. Letters indicate statistically significant differences between groups. [b]Total POS: ΣDP2, Di-GalA, DP3, Tri-GalA, and DP > 3.

pectins were carried out using 16 U mL^{-1} of Viscozyme even after 30 min of reaction, a high hydrolysis rate was found and most oligosaccharides (>80%) had an M_w below 1 kDa. The chromatographic profiles of hydrolyzed citrus and apple pectin using Viscozyme (4 U mL^{-1}) and Glucanex (16 U mL^{-1}) are shown in Figure 2. As it can be seen, the molecular mass average (M_w) distributions of hydrolysates were very different. Samples hydrolyzed with Glucanex (red line) showed a wider variety of fragments and higher M_w than those treated with Viscozyme (blue line). In contrast, chromatographic profiles of hydrolyzed citrus and apple pectin with Viscozyme were similar, despite differences in the monomer composition and the initial M_w of both pectins, 547 and 427 kDa for citrus and apple pectin, respectively, and 169 kDa for PGA. Similar profiles were also obtained when citrus and apple pectin were hydrolyzed with Glucanex.

Figure 3 shows the distribution of M_w of the different fragments (POS) produced and their concentrations (mg g^{-1} pectin) in hydrolysates at 30 min and 24 h using Viscozyme (4 U mL^{-1}) and Glucanex (16 U mL^{-1}). As already stated, similar hydrolysis patterns were obtained for the two pectins using Viscozyme, and the same behavior was observed when pectins were hydrolyzed with Glucanex. With respect to PGA, the behavior was similar when hydrolysis was performed using Glucanex, while PGA was more quickly hydrolyzed with Viscozyme.

The quantitative data of low M_w compounds released during pectin hydrolysis showed those differences between enzymes. After 30 min of incubation with Glucanex, POS with a M_w of around of 7−11 kDa comprised 36.3%, 20.7%, and 8.3% of total POS obtained from citrus and apple pectin and PGA, respectively, and fragments of around 2.6−3.8 kDa were approximately 30.1%, 53.9%, and 33.7% for the same samples. Also, POS with M_w from 0.8 to 0.9 and di- and monosaccharides (0.2−0.3 kDa) were determined in assayed pectins and PGA. After 24 h of hydrolysis, the release of low M_w (<1 kDa) compounds was very high.

In contrast, after 30 min of enzymatic hydrolysis with Viscozyme, most obtained POS had a M_w of 2.4−2.5 kDa, comprised of around 76.4% and 78.7% for citrus and apple pectin. After 24 h of reaction, >80% of the fragments were below 1 kDa in all hydrolysates.

These results corroborated the importance of enzyme type in the formation of POS. Several studies reported the influence of enzyme and reaction conditions over POS formation.[30,31] Pectin source is another influential factor for the type and yield of POS formed; thus, enzymatic hydrolysis of highly methylated and low methylated pectin by endopolygalactur-

DOI: 10.1021/acs.jafc.9b00930
J. Agric. Food Chem. 2019, 67, 7435−7447

onase (endo-PG) that randomly hydrolyses the homogalacturonan backbone yielded 95% POS with an average molecular mass of 3.5 and 3.0 kDa, respectively.[30] Thus, POS with a similar M_w may be obtained by using pectin from different sources.[32] A mixture of four endoenzymes (pectin methyl esterase, endo-PG, endoarabinosidase, and endogalactosidase) has been used in order to obtain POS of around 3 kDa from sugar beet pectin, predominantly 0.9 kDa oligogalacturonides.[33]

In order to know possible structures of POS present in hydrolysates of citrus and apple pectins using Glucanex or Viscozyme, HPAEC-PAD analysis was carried out; Figure 4 shows the chromatographic profiles of carbohydrates found. Acidic sugars, GalA (peak 1), Di-GalA (peak 2), and Tri-GalA (peak 3) eluted after neutral sugars. As with HPSEC-RID profiles of pectins, differences in POS profiles depending on the enzyme preparation were observed. In Glucanex (blue line) hydrolysates, a wide variety of acidic peaks with a degree of polymerization (DP) of 2, 3, and >3 were detected. When the retention times were compared with standards of GalA and di- and tri-GalA, these compounds could be considered oligosaccharides with one molecule of GalA and one or more molecules of neutral sugars. This different pattern could highlight a different behavior of these enzymes; thus, Viscozyme could have an important endo-PG activity while Glucanex could have other glycosidase activities. All POS formed were quantified (Table 2). It can be observed that Viscozyme releases higher amounts of Di-GalA and Tri-GalA than Glucanex, while this enzyme released a great amount of disaccharides with GalA. HPAEC-PAD profiles showed that high mono-, di-, and tri-GalA levels were detected. According to the results obtained by HPAEC-PAD, POS formation from apple and citrus pectin after 30 min of hydrolysis ranged from 335 to 652 mg g^{-1} pectin. The highest formation of POS (767 mg g^{-1}) was obtained with Viscozyme preparation using standard PGA as substrate.

Other authors reported the use of Viscozyme to obtain POS from different origins, either directly from pectin-rich by-products, such as lemon peel, orange peel, and sugar beet pulp,[3,4,34] or pectin isolated from sugar beet pulp, onion skins, rapeseed cake, and PGA.[5,6,30,35] Our results agree with these studies where high amounts of low M_w POS were obtained. On the other hand, Glucanex has been employed in the digestion of protoplasts from fungal spores[36] and to obtain POS and modified pectin from artichoke pectin.[7] In this previous study, a low concentration of enzyme was used (0.63 U mL^{-1}); modified pectin with a M_w of 61 kDa was obtained, determined by HPSEC-ELSD, and a very low amount of POS, 63.7 mg g^{-1}, was quantified by HPAEC-PAD.

Characterization of POS Present in Reaction Mixtures. To gain a deeper knowledge of the structure of POS released during enzymatic hydrolysis of apple and citrus pectins and PGA, GC-MS analyses of hydrolysates were carried out. Neutral monosaccharides, GalA, Di-GalA, and Tri-GalA as well as unknown di- and trisaccharides, perhaps corresponding to compounds eluted in HPAEC-PAD profiles between GalA and Di-GalA (Figure 4), were detected in the GC-MS profiles of citrus and apple pectin and PGA hydrolyzed with Glucanex (Figure S3a). Moreover, m/z ions 277, 321, 332, 333, 423, and 540 were present in the MS spectra of Di-GalA (Figure S3b). As previously indicated by Füzfai et al.,[37] m/z 332 and 540 were characteristic ions from GalA. In addition, m/z ions 277, 321, 333, and 423, derived from β-cleavage fragmentation of uronic acids, were reported in small abundance.[38]

GC-MS Spectra Classification Using Supervised Machine Learning Algorithms. Due to the scarce information found in the bibliography about the GC-MS spectra of these type of compounds, analysis of MS spectra employing machine learning algorithms (RF and BLR) was performed using full mass spectra of all disaccharides and trisaccharides found in the enzymatic hydrolysates, except Di- and Tri-GalA (identified with standards and their structure confirmed by GC-MS); Table S1 shows the number of spectra used to perform this study.

Before data analysis, each MS spectrum was decomposed and reconstructed using the DWT. Models were built using ions whose abundances were statistically different among groups ($p < 0.05$) and might correspond to known POS ruptures, assessed by competitive fragmentation modeling (CFM-ID),[14] as inputs.

Unknown POS were classified according to three different criteria: (i) sample type (apple pectin, $n = 40$; citrus pectin, $n = 48$; PGA, $n = 62$; known sugars, $n = 62$), (ii) enzyme type (Glucanex, $n = 70$; Viscozyme, $n = 80$; known sugars, $n = 62$), and (iii) sample and enzyme type (Glucanex-apple-POS, $n = 18$; Viscozyme-apple-POS, $n = 22$; Glucanex-citrus-POS, $n = 12$; Viscozyme-citrus-POS, $n = 36$; Glucanex-PGA-POS, $n = 40$; Viscozyme-PGA-POS, $n = 22$; known sugars, $n = 62$). All models were validated and tested on new samples.

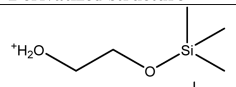
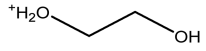
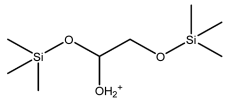
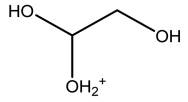
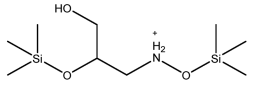
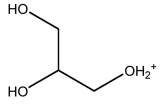
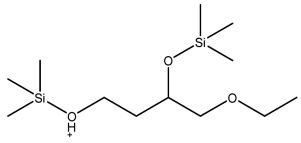
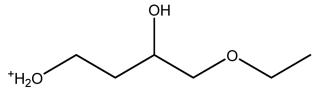
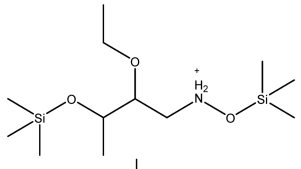
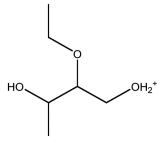
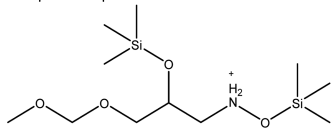
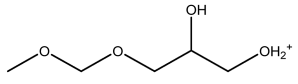
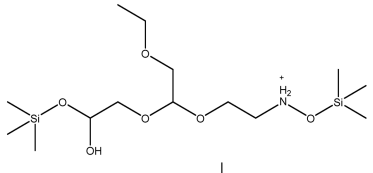
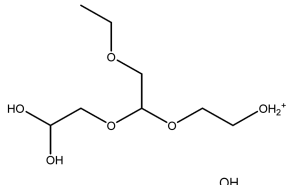
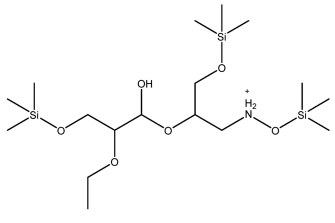
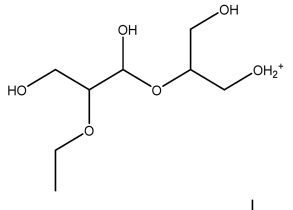
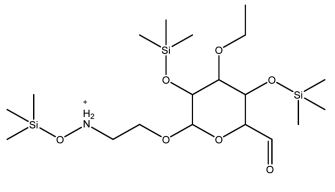
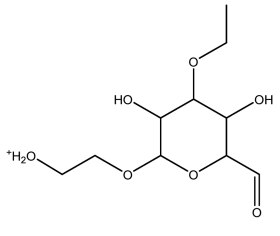
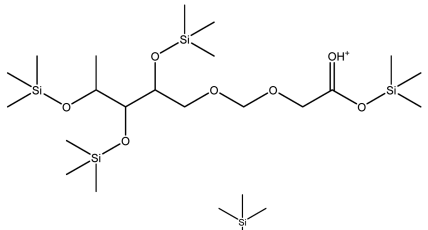
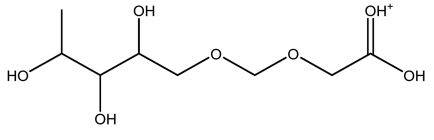
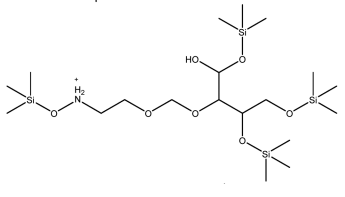
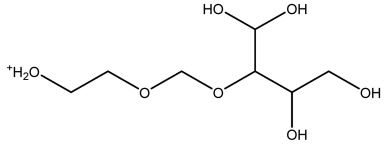
When interpreting these models, it should be considered that RF, BLR, and MLP are computed in a different way, leading to different results. These models may show a better/worse performance depending on the spectral data and its applications. RF builds multiple decision trees, outputting the different classes (i.e., enzyme used, pectin source), and each node is split using the best among a subset of predictors (i.e., m/z ions) randomly chosen. Then, RF averages the results from each tree to get a more accurate and stable prediction. BLR is considered as an ensemble method that uses a weighted average of predictions of individual classifiers. The term "boosting" refers to using a set of weak models, each slightly different, to build a strong model. In BLR, the base classifier (weak model) is logistic regression. The term iterations specify the maximal number of models to be fitted. On the other hand, MLP is the most common kind of ANN, one of the most popular families of machine learning models that allow one to model complex and highly nonlinear processes. MLP is formed by an input layer (i.e., m/z ions), an output layer (i.e., enzyme used, pectin source), and several neurons or nodes organized in hidden layers, where each neuron in a layer is connected with each neuron in the next layer through a weighted connection. The activation function is a transformation applied to the input spectra to determine whether the information that the neuron is receiving is relevant or not.

The training, cross-validation, and test rates for the sample type classification study were, respectively: RF, 100%, 98.6%, and 100%; BLR, 100%, 96.5%, and 100%.

Kappa values for RF and BLR were 0.98 and 0.95, respectively. Kappa is a robust measure considering the possibility of a correct classification by chance. Taking into account that no misclassifications occurred during the test phase, model sensitivity, specificity, balanced accuracy between sensitivity and specificity, and true positive and negative rates were 100%. The most influential m/z ions in each model were determined (Tables S2 and S3).

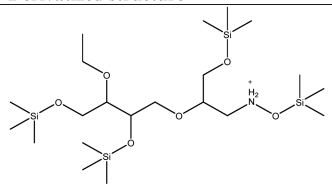
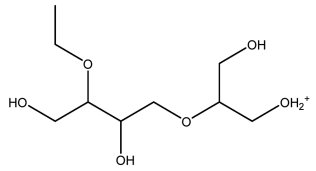
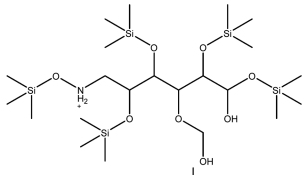
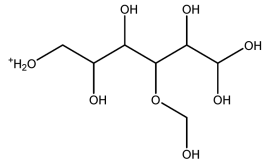
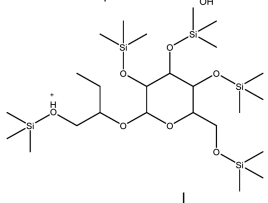
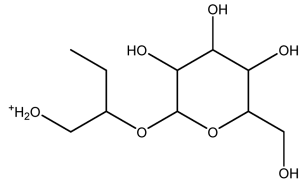
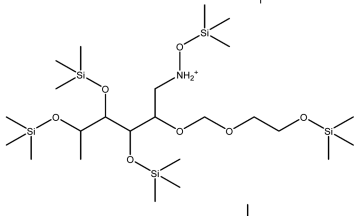
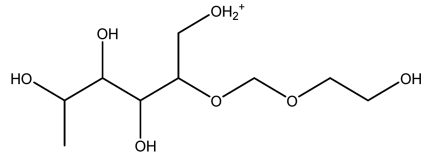
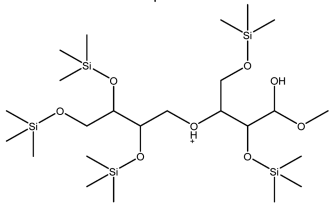
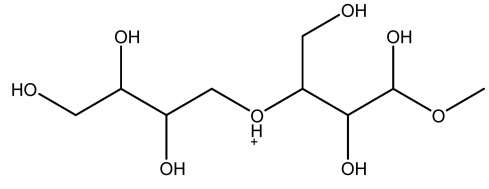
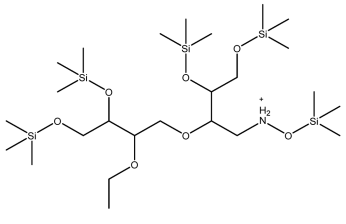
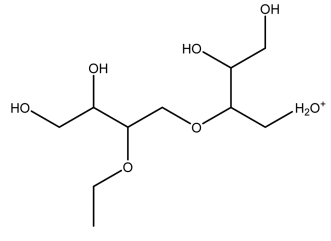
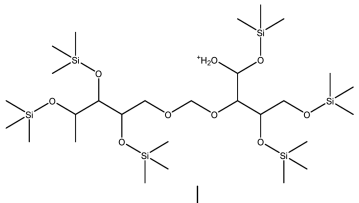
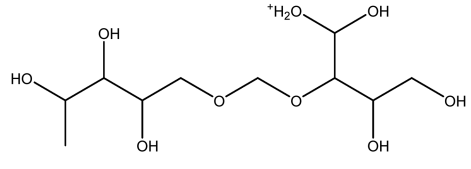
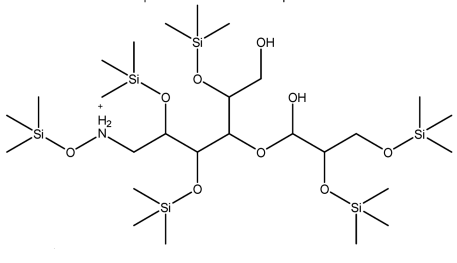
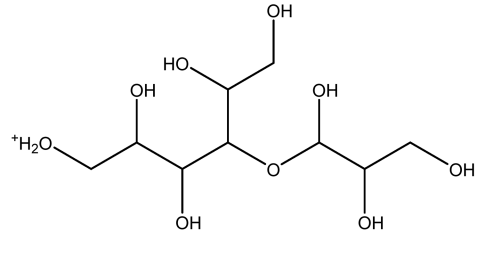
DOI: 10.1021/acs.jafc.9b00930
J. Agric. Food Chem. 2019, 67, 7435−7447

Table 3. Possible Chemical Structures of Some of the Most Influential *m/z* Ions Given by Machine Learning Algorithms, Determined by Competitive Fragment Modelling (CFM-ID), Selected for the 12 Rules[a]

m/z	Derivatized structure[b]	Original molecule fragment
135		
223		
252		
279		
294		
296		
370		
442		
482		
513		
516		

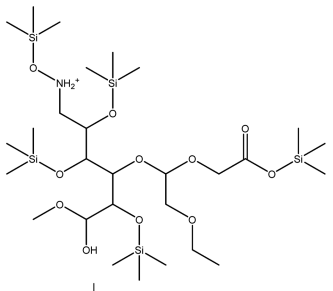
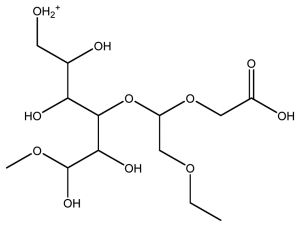
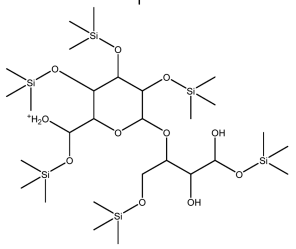
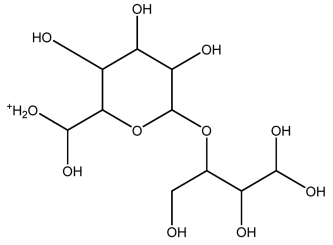
DOI: 10.1021/acs.jafc.9b00930
J. Agric. Food Chem. 2019, 67, 7435−7447

Table 3. continued

m/z	Derivatized structure[b]	Original molecule fragment
528		
604		
613		
616		
617		
630		
719		
720		

DOI: 10.1021/acs.jafc.9b00930
J. Agric. Food Chem. 2019, 67, 7435−7447

Table 3. continued

m/z	Derivatized structure[b]	Original molecule fragment
734		
749		

[a]They correspond to TMSO fragments from pectic di- or trisaccharides present in enzymatic hydrolysates of citrus and apple pectin and PGA.

Then, POS were classified according to their enzymatic origin with the following training, cross-validation, and test rates, respectively: RF, 100%, 98.6%, and 100%; BLR, 100%, 98.6%, and 100%.

Kappa values for RF and BLR were 0.98 for both cases. Again, no misclassifications were observed, so model sensitivity, specificity, balanced accuracy, and true positives and negatives rates were 100%. Also, the most influential m/z ions in each model were determined (Tables S4 and S5).

Finally, the predictive power of these algorithms was ensured by classifying POS according to both enzyme origin and sample type. In addition, RF and BLR performance was compared to a third model that also showed high prediction rates in this study, an artificial neural network, MLP. Training, cross-validation, and test rates were, respectively: RF, 100%, 97.3%, and 100%; BLR, 100%, 97.9%, and 100%; MLP: 100%, 88.7%, and 90.6%

Kappa values for RF, BLR, and MLP were 0.97, 0.97, and 0.86. RF and BLR showed significantly higher overall accuracy and kappa values than MLP (Figure S4), confirming their high prediction rates. For MLP, lower model sensitivities were obtained for Glucanex-PGA-POS and Glucanex-Apple-POS (73.3% and 85.7%, respectively) while model specificity was above 95% in all cases. The true positives rate was lower for Viscozyme-PGA-POS and known sugars (71.4% and 89.5%, respectively), and the true negatives rates were higher than 90%. Balanced accuracy ranged from 85.7% to 100% (Table S6). For RF and BLR, these rates were 100%. The most influential m/z ions in each model were determined (Tables S7, S8, and S9).

It has been suggested that RF is a more robust model compared to BLR and MLP, leading to a higher accuracy when tested on new samples (test rate). This fact was also reported in our previous work.[7] However, model performance depends on input data, so a comparison between algorithms was needed. Moreover, these algorithms reinforce each other and are different from conventional chemometric methods. PCA is not able to properly discriminate between structures obtained from specific substrates using a specific enzyme (Figure S5a) showing a poor performance (low percentages of variance explained by the first components). In contrast, RF, BLR, and MLP showed high classification accuracies when tested on new samples as explained above. To illustrate this high predictive power, the model performance was compared to the one obtained in a traditional supervised method like PLS-DA (Figure S6). It should be considered that oligosaccharide GC-MS-EI fragmentation is a complex process that produces high dimensional data that may exhibit a significant degree of nonlinearity. In addition, differences in GC-MS-EI spectra of similar molecules like POS structures are very subtle and may not be discriminated by conventional models. Therefore, advanced pattern recognition methods are needed.

We have demonstrated that it is possible to get highly accurate classifications of POS based on several mathematical approaches. The ability to classify POS highlights structural differences according to enzyme and pectin sources.

These models discriminated POS obtained with the same enzyme preparation and different pectin, although these hydrolysates gave very similar HPSEC-RID chromatographic profiles.

Selection of the Most Relevant Fragments Using Random Forests. Machine learning algorithms detected structural differences in POS GC-MS spectra according to their enzyme and sample types and were also able to discriminate between POS obtained from a specific substrate with a specific enzyme preparation with RF being a highly accurate method. To interpret these results, RF built was also used for selecting the most relevant m/z fragments in these classifications via backward feature elimination. This method is able to determine the variables that had the strongest impact on the classification and has been used to select a set of lipids as nutritional biomarkers in infant metabolism.[39]

Therefore, 12 fragments explained differences according to the sample type (m/z 121, 370, 514, 603, 616, 629, 646, 657, 699, 703, 733, and 747); 30 fragments explained differences according to the enzyme type (m/z 123, 135, 151, 235, 238, 252, 279, 294, 309, 310, 442, 458, 513, 516, 528, 546, 603, 604, 613, 616, 617, 630, 660, 704, 719, 731, 732, 733, 734, and 749), and 11 fragments explained differences according to both pectin and enzyme types (m/z 107, 121, 223, 252, 296, 482,

DOI: 10.1021/acs.jafc.9b00930
J. Agric. Food Chem. 2019, 67, 7435−7447

617, 646, 719, 720, and 734). Some of these fragments were influential in more than one classification, so a total of 44 m/z different fragments were considered.

Probable Structures of Relevant Fragments and Chemical Origin. Probable chemical structures of selected fragments have been suggested using CFM-ID code.[14] Proposed structures are shown in Tables 3 and S10. For this purpose, a POS *in silico* fragmentation library was built by calculating complete feasible fragments for candidate molecules (i.e., pectic monosaccharides, POS-DP2, and DP3), systematically breaking bonds within the molecule.[13,14] For monosaccharides, pectic neutral sugars (arabinose, rhamnose, xylose, and galactose) and GalA (methylated/acetylated or not) were considered. The POS-DP2 included, taking into account more possible structures present in HG and RG-I, were Di-GalA (methylated or not), xylose-α(1,3)-GalA (methylated/acetylated or not), GalA-α(1,2)-rhamnose (acetylated or not), rhamnose-α(1,4)-GalA (acetylated or not), galactose-α(1,4)-rhamnose, and arabinose-α(1,4)-rhamnose.[8,21] A Tri-POS library was generated, considering combinations of all POS-DP2 structures and pectic monomers, following the main criteria reported by those authors. The criteria were established following a previous work.[7] Then, we selected a chemical structure for each m/z ion by looking for these specific ruptures in our library. Spectral fragments, m/z, with statistically significant differences ($p < 0.05$) are reflected below.

Sample Type Study. Several m/z ions were more abundant in **known compounds** (pectic sugars and Di- and TriGalA). These ions include m/z 121, 370, 514, 603, 616, 646, 657, 699, 703, 733, and 747. All these fragments could be derived from GalA-α(1,4)-GalA oligomers. In addition, m/z 370 formed from GalA-containing POS was high in **PGA-POS**. On the other hand, the abundance of m/z 121, derived from all POS containing GalA-α(1,4)-GalA, was high in **apple-POS** while m/z 629 might be originated from GalA-α(1,2)-rhamnose, rhamnose-α(1,4)-GalA, and xylose-α(1,3)-GalA, and it was lower in **apple-POS**.

Enzyme Type Study. Like the previous study, ions m/z 294, 309, 310, 442, 458, 513, 516, 528, 546, 603, 604, 613, 616, 617, 630, 660, 704, 719, 731, 732, 733, and 734 showed higher abundances in **known sugars** than in unknown POS. These ions could be derived from GalA-containing POS. From them, ions m/z 309, 458, 513, 528, and 546 were previously reported by Sabater et al.,[7] showing high abundances in known pectic sugars too. Other ions (m/z 123, 135, 151, 235, 238, 252, and 279) were more abundant in **Viscozyme-POS**; from them, nonspecific ions m/z 135, 151, and 252 could be formed from all pectic sugars and POS ruptures while m/z 123 could be derived from POS containing methylated GalA. Other more specific ions are m/z 235 from POS containing neutral sugars (galactose-α(1,4)-rhamnose, arabinose-α(1,4)-rhamnose, GalA-α(1,2)-rhamnose, rhamnose-α(1,4)-GalA, and xylose-α(1,3)-GalA) and m/z 238 from acetylated GalA-α(1,2)-rhamnose and xylose-α(1,3)-GalA (acetylated), and m/z 279 might be specifically derived from acetylated GalA-α(1,2)-rhamnose. Finally, ion m/z 749, formed from GalA-containing POS, was high in **Glucanex-POS**.

Pectin and Enzyme Type Study. Ions m/z 617, 646, 719, 720, and 734 were high in **known pectic sugars**, and most of them were also characteristic for these compounds in the other comparative studies. In contrast, m/z 107, 121, 223, 252, 296, and 482 were high in **Viscozyme-apple-POS**; nonspecific ions m/z 107, 121, 223, and 252 are derived from GalA-containing POS while m/z 296 might be specifically originated from xylose-α(1,3)-GalA (methylated or not) rupture as well as from its two monomers, and m/z 482 was derived from GalA-α(1,2)-rhamnose (acetylated) and xylose-α(1,3)-GalA (acetylated), showing also high abundances for POS obtained with Cellulase from *A. niger* in our previous work.[7]

These results highlight structural differences between POS and indicate that it is possible to classify the oligosaccharides formed according to the enzyme used for their obtainment and pectin source. These models could be applied on new reaction mixtures with novel oligosaccharides to determine the pectin used as raw material or study the similarities of POS obtained with different commercial enzyme preparations considering the high prediction rates on new samples.

Association Rules between Relevant Fragments. To ensure that these models extract the most relevant structural information from GC-MS spectra, selected ions in RF studies (Table 3) were used to establish 12 association rules. When compared to simpler models, RF and other rule-based models provide additional information apart from variable importance analysis. RF generates classification rules based on m/z ion abundances learned directly from the input data (POS GC-MS-EI spectra) (Figure S7). These rules indicate possible relationships between ions originated from specific POS structures, and possible parent molecules, present in each hydrolysate, could be suggested. On the other hand, no clear relationships between specific ions related to specific hydrolysates can be proposed from the loadings plot of simple models like PCA (Figure S5b). However, the RF model consists of 1000 different trees, and relevant rules between m/z ions should be extracted to deepen model interpretation. To summarize the most important relationships between POS m/z ions, association rules were calculated. Association rule modeling involves fewer rules and contains contextual information within the ruleset, so they can be easily interpreted. Moreover, the calculated association rules showed confidence values of 1.0; this fact means that ions described in each rule are formed together in 100% of cases, so these rules are highly reproducible and could be generalized and applied on similar oligosaccharide structures or POS from other sources.

These rules (Figure S8) indicate which m/z ions originated from the same POS structures, considering both pectin and enzyme types. The first 6 rules were more characteristic of known pectic sugars.

Rule 1: m/z 135, 223, and 516 in known sugars. These three ions are formed from the rupture of GalA-α(1,4)-GalA units.

Rule 2: m/z 252, 294, 370, and 516 in known sugars. Similar to rule 1, all these ions correspond to different ruptures of GalA dimers and could be derived from m/z 516.

Rule 3: m/z 296, 370, and 613 in known sugars (probably DiGalA and TriGalA). Ion m/z 296 is derived from xylose or GalA ruptures and dimers of these two sugars and is originated with m/z 370 and 613, formed from GalA dimers or trimers bonded by α(1,4) glycosidic linkages.

Rule 4: m/z 296, 370, and 513 in known sugars.

Rule 5: m/z 296 and 734 in known sugars.

Rule 6: m/z 296 and 720 in known sugars.

As previously commented, m/z 296 corresponds to xylose or GalA ruptures while the other ions in rules 4, 5, and 6 are formed from GalA dimers or trimers. On the other hand, the

DOI: 10.1021/acs.jafc.9b00930
J. Agric. Food Chem. 2019, 67, 7435−7447

Table 4. Most Probable Units Found in Pectic Oligosaccharides (POS) Structures Produced from Polygalacturonic Acid (PGA), Highly Methyl-Esterified Pectins (Citrus and Apple), and Low Methyl-Esterified Pectin (Artichoke) by Enzymatic Hydrolysis Using Different Enzyme Preparations[a]

			POS containing GalA and Rha units			
Substrate	Enzyme	Article	GalA-α(1,2)-Rha	GalA-α(1,2)-Rha (acetylated)	Rha-α(1,4)-GalA	Rha-α(1,4)-GalA (acetylated)
PGA	Viscozyme (endo-PG)	Present study	Yes	Yes	-	-
PGA	Glucanex (exo-Gal)	Present study	-	-	-	-
Citrus	Viscozyme (endo-PG)	Present study	Yes	Yes	-	-
Citrus	Glucanex (exo-Gal)	Present study	-	-	-	-
Apple	Viscozyme (endo-PG)	Present study	-	Yes	-	-
Apple	Glucanex (exo-Gal)	Present study	-	Yes	-	-
Artichoke	Glucanex (exo-Gal)	Sabater et al. (2019)	-	Yes	-	Yes
Artichoke	Pentopan (1,4-endoxylanase)	Sabater et al. (2019)	Yes	Yes	-	Yes
Artichoke	CAN (Cellulase)	Sabater et al. (2019)	-	Yes	-	-
Artichoke	PUO (Pectin-lyase)	Sabater et al. (2019)	Yes	Yes	Yes	Yes

			POS containing GalA and Xy units			POS containing Ara and Rha units
			Xyl-α(1,3)-GalA	Xyl-α(1,3)-GalA (methylated)	Xyl-α(1,3)-GalA (acetylated)	Ara-α(1,4)-Rha
PGA	Viscozyme (endo-PG)	Present study	-	-	Yes	-
PGA	Glucanex (exo-Gal)	Present study	Yes	Yes	Yes	-
Citrus	Viscozyme (endo-PG)	Present study	-	-	Yes	-
Citrus	Glucanex (exo-Gal)	Present study	-	-	-	-
Apple	Viscozyme (endo-PG)	Present study	Yes	Yes	Yes	-
Apple	Glucanex (exo-Gal)	Present study	-	-	-	-
Artichoke	Glucanex (exo-Gal)	Sabater et al. (2019)	-	-	Yes	-
Artichoke	Pentopan (1,4-endoxylanase)	Sabater et al. (2019)	Yes	Yes	Yes	Yes
Artichoke	CAN (Cellulase)	Sabater et al. (2019)	-	-	Yes	-
Artichoke	PUO (Pectin-lyase)	Sabater et al. (2019)	Yes	Yes	Yes	-

[a]Xyl: xylose; Ara: arabinose; Rha: rhamnose; GalA: galacturonic acid; DM: degree of methyl-esterification; PG: polygalacturonase; exo-Gal: exo-β-D-galactofuranosidase; CAN: Cellulase from *Aspergillus niger*; PUO: Pectinex Ultra Olio.

rest of the rules were more characteristic of unknown POS found in hydrolysates.

Rule 7: m/z 279 and 528 in Glucanex-apple-POS. Ion m/z 279 derives from acetylated GalA-α(1,2)-rhamnose while m/z 528 is formed from GalA dimers. These results could indicate different ruptures of oligosaccharides containing one/two units of GalA bonded to one rhamnose molecule.

Rule 8: m/z 296, 516, and 719 in Glucanex-PGA-POS. These ions are formed from GalA chains and may also be from POS containing one/two units of GalA bonded to one xylose molecule.

Rule 9: m/z 616 and 749 in Viscozyme-citrus-POS.

Rule 10: m/z 604 and 749 in Viscozyme-apple-POS.

Rule 11: m/z 442 and 617 in Viscozyme-PGA-POS.

All ions from rules 9, 10, and 11 are formed from GalA-α(1,4)-GalA units indicating HG ruptures by Viscozyme preparation.

Rule 12: m/z 482 and 630 in Viscozyme-apple-POS. Ion m/z 482 is derived from GalA-α(1,2)-rhamnose (acetylated) or xylose-α(1,3)-GalA (acetylated) while m/z 630 is formed from GalA chains, indicating the rupture of larger structures containing several units of GalA with rhamnose or xylose residues.

As indicated above, association rules determine which ions are formed together, and possible parent molecules may be suggested (Figure S8). Considering all ions associated by each rule are originated together, these ions share common regions in their structure and derive from larger oligosaccharide structures (Figure S9). To indicate potential parent molecules, it should be noted that these ions derive from specific POS according to our *in silico fragmentation* library. Ions from rules 1−6 derive from galacturonic acid (GalA) oligomers and are associated, indicating the presence of DiGalA and TriGalA. Fragments from rule 7 are formed from acetylated GalA-α(1,2)-rhamnose and GalA dimers, indicating the presence of acetylated DiGalA-α(1,2)-rhamnose. Ions from rule 8 derive from GalA oligomers and xylose-α(1,3)-GalA units, so xylose-α(1,3)-DiGalA is present in these hydrolysates. Fragments from rules 9−11 are relevant in unknown POS (not DiGal nor TriGalA) and derive from GalA-α(1,4)-GalA units, so the presence of trisaccharides consisting of GalA-α(1,4)-GalA attached to neutral sugars can be suggested. Finally, ions from rule 12 are formed from GalA-α(1,4)-GalA, GalA-α(1,2)-rhamnose (acetylated), or xylose-α(1,3)-GalA (acetylated) units, indicating the presence of acetylated DiGalA-α(1,2)-rhamnose or acetylated xylose-α(1,3)-DiGalA. It should be noted that these POS structures could not have been elucidated on the basis of variable importance analysis of traditional methods.

Statistical differences among relevant ions and association analysis give complementary information. It can be suggested that both Viscozyme and Glucanex produce POS containing several units of GalA and one unit of rhamnose or xylose. To verify the applicability of these rules, GC-MS spectra were classified according to both, pectin and enzyme types, in a similar way to the third machine learning study (iii) but considering only these rules (not m/z abundances). Train and test rates were 100% and 95.3%, respectively, with a kappa value of 0.94. Model sensitivity was low for Glucanex-citrus-POS (66.7%), but model specificity was above 98% for all the classes. True positives rate was low in Glucanex-apple-POS and Viscozyme-apple-POS (83.3% and 85.7%) and Viscozyme-PGA-POS (85.7%), but true negative rate was high in all cases.

DOI: 10.1021/acs.jafc.9b00930
J. Agric. Food Chem. 2019, 67, 7435−7447

Balanced accuracies ranged from 83.3% to 100%, showing the lowest values for Glucanex-citrus-POS (Table S11).

There is a wide variety of novel pectin sources and enzymes that can be used for POS obtainment. This makes it difficult to predict POS from a specific source. However, the present study and our previous study,[7] reporting depolymerization of low methyl-esterification pectin from artichoke (DM 19.5 ± 0.0%), allow us to tentatively predict some of the most probable POS structures that may be obtained from HG type polymer (PGA) as well as high or low methyl-esterification pectin using different enzymes. These structures are reported in Table 4. Interestingly, production of POS containing rhamnose and GalA dimers was correlated to the pectin content of these two sugars as well as the DM of pectin. Rhamnose-α(1,4)-GalA (acetylated or not) was produced only from low methyl-esterification pectin with high neutral sugar content, while POS containing units of GalA-α(1,2)-rhamnose were also produced from apple and citrus pectin when Viscozyme (an endo-PG) was used. Similarly, production of POS containing xylose and GalA was correlated to the DM. Xylose-α(1,3)-GalA containing POS were produced mainly from PGA (due probably to the presence of xylogalacturonan) as well as apple pectin and artichoke pectin. Arabinose containing POS are mainly produced from low methyl-esterification pectin with high neutral sugar content using only a xylanase that disrupts the RG-I branch.

In conclusion, using traditional techniques, it was observed that the same enzyme preparation applied on different pectin sources gave rise to hydrolysates with similar chromatographic profiles. In contrast, these two enzymatic preparations produced POS with different profiles regardless on pectin source. Considering the HPSEC-RID results, Glucanex produced POS with a higher M_w (around 7 kDa) than Viscozyme, (around 2.4 kDa). Moreover, when POS of low M_w were analyzed by HPAEC-PAD and GC-MS, a higher amount of Di-GalA and Tri-GalA was formed with Viscozyme, showing a different profile of released neutral and acid carbohydrates. In regard to the GC-MS characterization of the POS structures, this work should be considered as a first approach to a comprehensive study. Enzymatic depolymerization of two common highly methyl-esterified pectins (from citrus and apple) as well as PGA, used as HG standard, employing two enzymes with very different activities was studied. Thus, POS characterization using a relatively simple technique like GC-MS-EI allows one to verify structural similarities and differences in POS obtained from the most widely used pectin sources. Mass spectra of di- and trisaccharides derived from GalA have been obtained from apple and citrus pectin and PGA for the first time. In addition, characteristic GC-MS-EI patterns were attributed to the presence of specific units in POS structures. However, only small oligosaccharide molecules are detected by GC-MS. Dimers and trimers described in this study are also present in larger molecules produced during enzymatic hydrolysis. To characterize larger structures, more studies dealing with more advanced MS-based techniques are needed.

The results presented here improve existing computational methods by extracting more information from oligosaccharide GC-MS-EI spectra. Finally, this methodology could be applied to novel reaction mixtures or oligosaccharides from unknown origin obtained with different enzymes to determine similar hydrolysis patterns and find possible structure−function relationships.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jafc.9b00930.

(i) Pectin depolymerization experiments and GC-FID/GC-MS chromatographic profiles (Figures S1−S3, Table S1), (ii) model performance and comparison of different algorithms studied for spectra classification (Figures S4−S7, Tables S2−S9 and S11), (iii) chemical structures of relevant m/z ions present in specific POS spectra (Table S10), and (iv) association rules between m/z ions and common substructures of fragments formed together (Figures S8 and S9) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Tel: +34 910017952. E-mail: a.montilla@csic.es.

ORCID

Antonia Montilla: 0000-0003-2117-6999

Nieves Corzo: 0000-0002-6420-2344

Funding

This work has been funded by MICINN of Spain, Projects AGL2014-53445-R and AGL2017-84614-C2-1-R. C.S. is thankful for his FPU Predoc contract from Spanish MECD (FPU14/03619).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Ramiro Martínez (Novozyme) for the enzyme supply.

■ REFERENCES

(1) de Moura, F. A.; Macagnan, F. T.; da Silva, L. P. Oligosaccharide production by hydrolysis of polysaccharides: a review. *Int. J. Food Sci. Technol.* **2015**, *50*, 275−281.

(2) Holck, J.; Hotchkiss, A. T.; Meyer, A. S.; Mikkelsen, J. D.; Rastall, R. A. Production and bioactivity of pectic oligosaccharides from fruit and vegetable biomass. In *Food Oligosaccharides. Production, Analysis and Bioactivity*; Moreno, F. J., Sanz, M. L., Eds.; John Wiley & Sons, Ltd.: Chichester, UK, 2014; pp 76−87.

(3) Gómez, B.; Gullón, B.; Yáñez, R.; Parajó, J. C.; Alonso, J. L. Pectic oligosacharides from lemon peel wastes: Production, purification, and chemical characterization. *J. Agric. Food Chem.* **2013**, *61*, 10043−10053.

(4) Martínez, M.; Gullon, B.; Schols, H. A.; Alonso, J. L.; Parajo, J. C. Assessment of the Production of Oligomeric Compounds from Sugar Beet Pulp. *Ind. Eng. Chem. Res.* **2009**, *48*, 4681−4687.

(5) Baldassarre, S.; Babbar, N.; Van Roy, S.; Dejonghe, W.; Maesen, M.; Sforza, S.; Elst, K. Continuous production of pectic oligosaccharides from onion skins with an enzyme membrane reactor. *Food Chem.* **2018**, *267*, 101−110.

(6) Cobs-Rosas, M.; Concha-Olmos, J.; Weinstein-Oppenheimer, C.; Zuniga-Hansen, M. E. Assessment of antiproliferative activity of pectic substances obtained by different extraction methods from rapeseed cake on cancer cell lines. *Carbohydr. Polym.* **2015**, *117*, 923−932.

(7) Sabater, C.; Olano, A.; Corzo, N.; Montilla, A. GC−MS characterisation of novel artichoke (*Cynara scolymus*) pectic-oligosaccharides mixtures by the application of machine learning algorithms and competitive fragmentation modelling. *Carbohydr. Polym.* **2019**, *205*, 513−523.

(8) Caffall, K. H.; Mohnen, D. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydr. Res.* **2009**, *344*, 1879−1900.
(9) Delattre, C.; Michaud, P.; Vijayalakshmi, M. A. New monolithic enzymatic micro-reactor for the fast production and purification of oligogalacturonides. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2008**, *861*, 203−208.
(10) Ma, X.; Wang, D.; Chen, W.; Ismail, B. B.; Wang, W.; Lv, R.; Ding, T.; Ye, X.; Liu, D. Effects of ultrasound pretreatment on the enzymolysis of pectin: Kinetic study, structural characteristics and anti-cancer activity of the Hydrolysates. *Food Hydrocolloids* **2018**, *79*, 90−99.
(11) Muñoz-Almagro, N.; Montilla, A.; Moreno, F. J.; Villamiel, M. Modification of citrus and apple pectin by power ultrasound: Effects of acid and enzymatic treatment. *Ultrason. Sonochem.* **2017**, *38*, 807−819.
(12) Leijdekkers, A. G.; Huang, J. H.; Bakx, E. J.; Gruppen, H.; Schols, H. A. Identification of novel isomeric pectic oligosaccharides using hydrophilic interaction chromatography coupled to traveling-wave ion mobility mass spectrometry. *Carbohydr. Res.* **2015**, *404*, 1−8.
(13) Allen, F.; Greiner, R.; Wishart, D. Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification. *Metabolomics* **2015**, *11*, 98−110.
(14) Allen, F.; Pon, A.; Greiner, R.; Wishart, D. Computational prediction of electron ionization mass spectra to assist in GC/MS compound identification. *Anal. Chem.* **2016**, *88*, 7689−7697.
(15) Yi, L.; Dong, N.; Yun, Y.; Deng, B.; Ren, D.; Liu, S.; Liang, Y. Chemometric methods in data processing of mass spectrometry-based metabolomics: a review. *Anal. Chim. Acta* **2016**, *914*, 17−34.
(16) Lim, D. K.; Long, N. P.; Mo, C.; Dong, Z.; Cui, L.; Kim, G.; Kwon, S. W. Combination of mass spectrometry-based targeted lipidomics and supervised machine learning algorithms in detecting adulterated admixtures of white rice. *Food Res. Int.* **2017**, *100*, 814−821.
(17) Moncayo, S.; Manzoor, S.; Rosales, J. D.; Anzano, J.; Caceres, J. O. Qualitative and quantitative analysis of milk for the detection of adulteration by Laser Induced Breakdown Spectroscopy (LIBS). *Food Chem.* **2017**, *232*, 322−328.
(18) Sabater, C.; Montilla, A.; Ovejero, A.; Prodanov, M.; Olano, A.; Corzo, N. Furosine and HMF determination in prebiotic-supplemented infant formula from Spanish market. *J. Food Compos. Anal.* **2018**, *66*, 65−73.
(19) Gosav, S.; Praisler, M.; Birsa, M. L. Principal component analysis coupled with artificial neural networks—A combined technique classifying small molecular structures using a concatenated spectral database. *Int. J. Mol. Sci.* **2011**, *12*, 6668−6684.
(20) Sabater, C.; Corzo, N.; Olano, A.; Montilla, A. Enzymatic extraction of pectin from artichoke (*Cynara scolymus L.*) by-products using Celluclast 1.5L. *Carbohydr. Polym.* **2018**, *190*, 43−49.
(21) Atmodjo, M. A.; Hao, Z. Y.; Mohnen, D. Evolving views of pectin biosynthesis. Ed. S.S. Merchant. *Annu. Rev. Plant Biol.* **2013**, *64*, 747−779.
(22) Aldrich, E. Wavelets: Functions for Computing Wavelet Filters, Wavelet Transforms and Multiresolution Analyses. *R package version 0.3−0*; 2013; https://CRAN.R-project.org/package=wavelets. Last accessed: 16/12/18.
(23) Bergmeir, C.; Benitez, J. M. Neural Networks in R Using the Stuttgart Neural Network Simulator: RSNNS. *J. Stat. Softw.* **2012**, *46*, 1−26.
(24) Liaw, A.; Wiener, M. Classification and Regression by Random Forest. *R News* **2002**, *2*, 18−22.
(25) Tuszynski, J. caTools: Tools: moving window statistics, GIF, Base64, ROC AUC, etc. *R package version 1.17.1*; 2014; https://CRAN.R-project.org/package=caTools. Last accessed: 16/12/18.
(26) Johnson, I.; Hahsler, M. arulesCBA: Classification Based on Association Rules. *R package version 1.1.3−1*; 2018; https://CRAN.R-project.org/package=arulesCBA. Last accessed: 16/12/2018.
(27) Soria, A. C.; Rodriguez-Sanchez, S.; Sanz, J.; Martinez-Castro, I. Gas Chromatographic Analysis of Food Bioactive Oligosaccharides. In *Food Oligosaccharides. Production, Analysis and Bioactivity*; Moreno, F. J., Sanz, M. L., Eds.; John Wiley & Sons, Ltd.: Chichester, UK, 2014; pp 370−398.
(28) Gama, R.; Van Dyk, J. S.; Pletschke, B. I. Optimisation of enzymatic hydrolysis of apple pomace for production of biofuel and biorefinery chemicals using commercial enzymes. *3 Biotechnol.* **2015**, *5*, 1075−1087.
(29) Robert, C.; Emaga, T. H.; Wathelet, B.; Paquot, M. Effect of variety and harvest date on pectin extracted from chicory roots (*Cichorium intybus* L.). *Food Chem.* **2008**, *108*, 1008−1018.
(30) Combo, A. M. M.; Aguedo, M.; Goffin, D.; Wathelet, B.; Paquot, M. Enzymatic production of pectic oligosaccharides from polygalacturonic acid with commercial pectinase preparations. *Food Bioprod. Process.* **2012**, *90*, 588−596.
(31) Babbar, N.; Baldassarre, S.; Maesen, M.; Prandi, B.; Dejonghe, W.; Sforza, S.; Elst, K. Enzymatic production of pectic oligosaccharides from onion skins. *Carbohydr. Polym.* **2016**, *146*, 245−252.
(32) Olano-Martin, E.; Mountzouris, K. C.; Gibson, G. R.; Rastall, R. A. Continuous production of pectic oligosaccharides in an enzyme membrane reactor. *J. Food Sci.* **2001**, *66*, 966−971.
(33) Ralet, M. C.; Cabrera, J. C.; Bonnin, E.; Quemener, B.; Hellin, P.; Thibault, J. F. Mapping sugar beet pectin acetylation pattern. *Phytochemistry* **2005**, *66*, 1832−1843.
(34) Martinez Sabajanes, M.; Yáñez, R.; Alonso, J. L.; Parajó, J. C. Pectic oligosaccharides production from orange peel waste by enzymatic hydrolysis. *Int. J. Food Sci. Technol.* **2012**, *47*, 747−754.
(35) Elst, K.; Babbar, N.; Van Roy, S.; Baldassarre, S.; Dejonghe, W.; Maesen, M.; Sforza, S. Continuous production of pectic oligosaccharides from sugar beet pulp in a cross flow continuous enzyme membrane reactor. *Bioprocess Biosyst. Eng.* **2018**, *41*, 1717−1729.
(36) Cheng, Y.; Bélanger, R. R. Protoplast preparation and regeneration from spores of the biocontrol fungus Pseudozyma flocculosa. *FEMS Microbiol. Lett.* **2000**, *190*, 287−291.
(37) Füzfai, Z.; Kovács, E.; Molnár-Perl, I. Identification and quantitation of the main constituents of sour cherries: Simultaneously, as their trimethylsilyl derivatives, by gas chromatography-mass spectrometry. *Chromatographia* **2004**, *60*, S143−S151.
(38) Peterson, G. Gas-chromatographic analysis of sugars and related hydroxy acids as acyclic oxime and ester trimethylsilyl derivatives. *Carbohydr. Res.* **1974**, *33*, 47−61.
(39) Acharjee, A.; Prentice, P.; Acerini, C.; Smith, J.; Hughes, I. A.; Ong, K.; Griffin, J. L.; Dunger, D.; Koulman, A. The translation of lipid profiles to nutritional biomarkers in the study of infant metabolism. *Metabolomics* **2017**, *13*, 25.

DOI: 10.1021/acs.jafc.9b00930
J. Agric. Food Chem. 2019, 67, 7435−7447

Annex III:

Other published articles

Food Chemistry 194 (2016) 6–11

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Short communication

Quantification of prebiotics in commercial infant formulas

Carlos Sabater, Marin Prodanov, Agustín Olano, Nieves Corzo *, Antonia Montilla

Instituto de Investigación en Ciencias de la Alimentación CIAL, (CSIC-UAM) CEI (UAM + CSIC), C/Nicolás Cabrera, 9, E-28049 Madrid, Spain

ARTICLE INFO

Article history:
Received 13 February 2015
Received in revised form 23 July 2015
Accepted 25 July 2015
Available online 29 July 2015

Keywords:
Infant formula
Prebiotics
GOS
FOS
Maltodextrins

ABSTRACT

Since breastfeeding is not always possible, infant formulas (IFs) are supplemented with prebiotic oligosaccharides, such as galactooligosaccharides (GOS) and/or fructooligosaccharides (FOS) to exert similar effects to those of the breast milk. Nowadays, a great number of infant formulas enriched with prebiotics are disposal in the market, however there are scarce data about their composition. In this study, the combined use of two chromatographic methods (GC-FID and HPLC-RID) for the quantification of carbohydrates present in commercial infant formulas have been used. According to the results obtained by GC-FID for products containing prebiotics, the content of FOS, GOS and GOS/FOS was in the ranges of 1.6–5.0, 1.7–3.2, and 0.08–0.25/2.3–3.8 g/100 g of product, respectively. HPLC-RID analysis allowed quantification of maltodextrins with degree of polymerization (DP) up to 19. The methodology proposed here may be used for routine quality control of infant formula and other food ingredients containing prebiotics.

1. Introduction

Although human milk is considered the best food to satisfy the nutritional needs of newborn, there are some situations where breastfeeding must be interrupted and infants are fed with infant formula (IF) or cases where mothers cannot produce enough milk to supply all of their baby's nutritional needs and they combine breast and IF feeding. The knowledge gained on chemical composition and biological properties of human milk allows adapting the composition of IFs to meet nutritional needs of newborn.

Human milk contains 8–13 g/L of a complex mixture of oligosaccharides which is about 20-fold higher than those of bovine milk (Urashima, Taufik, Fukuda, & Asakuma, 2013). This is one of the factors that explain, the higher level of intestinal bifidobacteria and lactobacilli, as well as the lower incidence of bacterial infections found in breast-fed infants (Barile & Rastall, 2013).

Since Gibson and Roberfroid (1995) introduced the concept of prebiotics as "non-digestible oligosaccharides that reach the colon without being hydrolysed and are selectively metabolized by health-positive bacteria such as bifidobacteria and lactobacilli thereby exerting a beneficial effect on the host health", extensive research has been carried out to identify prebiotic components. A growing number of *in vitro* and *in vivo* studies also show that prebiotics could induce beneficial physiological effects in the colon and also in extra-intestinal compartments or contribute towards the reduction of dysbiosis risk and associated intestinal and systemic pathologies (Roberfroid et al., 2010). Several other studies strongly suggest that human milk oligosaccharides (HMOs) act as prebiotics (Bode, 2009; Engfer, Stahl, Finke, Sawatzki, & Daniel, 2000) and have a wide range of biological activities. However, as HMOs are very complex glycans, their production at industrial scale is very difficult (Bode, 2009).

Only the enzymatic synthesis of some prebiotic such as galactooligosaccharides (GOS) and fructooligosaccharides (FOS) is a feasible alternative to produce them in the food industry. Particularly, it has been shown that the addition of different amounts of GOS, FOS or GOS/FOS mixtures to IF stimulates the growth of bifidobacteria and lactobacilli (Ben et al., 2008; Boehm et al., 2002; Moro et al., 2002), produces changes in the short chain fatty acids, making the profile of these acids closer to that observed in breast-fed infants (Knol et al., 2005), improves the stool characteristics (frequency, pH and softening) (Ben et al., 2008; Fanaro et al., 2005; Moro et al., 2002) and reduces the incidence of allergic manifestations and infections during the first two years of life (Arslanoglu et al., 2008).

IFs are composed basically by carbohydrates (54–61 g/100 g of product) and proteins (11–15 g/100 g of product). Depending on the type, sugars such as lactose, corn syrup, sucrose or starches have been successfully used as a source of carbohydrates (Morales, Olano, & Corzo, 2004). Besides, in the last years, prebiotic oligosaccharides have been added to IFs to mimic the benefits attributed to HMOs (Barile & Rastall, 2013; Braegger et al., 2011; Cilla, Lacomba, Garcia-Llatas, & Alegria, 2012). Despite the increasing use of prebiotics in IF production, there are scarce data about their prebiotic composition. In this work, a study on the carbohydrate composition of 24 commercial IFs, selected as representative of the Spanish market, has been carried out. Special attention has been paid to the determination of prebiotic carbohydrates.

* Corresponding author.
E-mail address: nieves.corzo@csic.es (N. Corzo).

http://dx.doi.org/10.1016/j.foodchem.2015.07.127
0308-8146/

2. Materials and methods

2.1. Reference substances and samples

Analytical reference substances such as fructose, galactose, glucose, *myo*-inositol, lactose, maltulose, maltose, kestose, nystose and maltodextrins with a degree of polymerization (DP) from 2 to 5 were purchased from Sigma (St. Louis, MO, USA). Raftilose® and Wako® FOS were from Orafti (Orafti, Barcelona, Spain) and Wako (Chemical Industries, Osaka, Japan), respectively. Vivinal® GOS were from Domo (Friesland Campina Domo, Amersfoort, The Netherlands).

Twenty-four IFs were purchased from several Spanish chemist's, corresponding to starting and follow-up formulas without prebiotics ($n = 8$), 4 with lactose (C) and 4 lactose free (LF), and prebiotic-enriched IF ($n = 16$), 3 with FOS (PFOS), 7 with GOS (PGOS), and 6 with mixtures of GOS/FOS (PGOS/FOS). Table 1 shows their carbohydrate composition as indicated on the product labels. All samples were analysed before their expiry date and all determinations were done in duplicate.

2.2. Determination of carbohydrates

Before chromatographic analysis, fat and protein interfering materials were removed by precipitation, using Carrez reagents (Moreno, Olano, Santa-Maria, & Corzo, 1999).

Carbohydrate (monosaccharides, disaccharides, GOS and FOS) quantification was carried out by GC-FID, following the method of Montilla, van de Lagemaat, Olano, and del Castillo (2006). Analysis of maltodextrins were performed by HPLC-RID according to Corzo-Martínez, Copoví, Olano, Moreno, and Montilla (2013).

To quantify GOS in IFs containing maltodextrins and to avoid interference in the GC analysis, samples were incubated at 37 °C for 24 h with α-amyloglucosidase (*Megazyme*©3300, Bray Co.

Table 2
Repeatability of the used GC method for determination of carbohydrates in a commercial infant formula containing prebiotics, prepared four times and analysed on the same day and on different days (4 days).

Carbohydrate	Concentration (mg/100 g of product)			
	Same day (four replicates)		Daily (during 4 days) (one replicate)	
	Average value	RSD* (%)	Average value	RSD (%)
Fructose	0.33	4.4	0.33	4.6
Galactose	0.02	8.3	0.02	8.8
Glucose	0.18	4.7	0.19	5.3
Myo-inositol	0.06	4.3	0.05	4.5
Sucrose	0.18	5.6	0.17	4.4
Lactose	41.6	3.9	41.7	3.6
Maltose	0.30	7.3	0.32	6.8
Trisaccharides	0.60	6.3	0.59	7.3
Tetrasaccharides	0.59	7.3	0.56	7.6
Pentasaccharides	0.25	6.9	0.27	7.4
Hexa and heptasaccharides	0.26	9.9	0.23	12.3

* RSD: Relative standard deviation.

Table 1
Carbohydrate composition of infant formulas available at the Spanish market, according to their package labels.

Product code	Total carbohydrates	Lactose	Prebiotics	Maltodextrins	Inositol	Observations
	(g/100 g product)				(mg/100 g)	
Infant formula without prebiotics						
Conventional (C)						
C1	57.8	57.8			80.0	
C2	61.7	–[a]		–	–	mdx
C3	60.0	–		12.3	–	Starch
C4	55.6	–		–	9.9	mdx
Lactose free (LF)						
LF1	58.6			–	36.0	mdx
LF2	60.8			60.8	9.9	mdx
LF3	57.6			36.1	47.0	mdx
LF4	55.0			–	30.0	mdx
Infant formula with prebiotics						
With fructooligosaccharides (PFOS)						
PFOS1	53.3	–	5.7	–	45.0	FOS/inulin
PFOS2	56.0	–	3.0	–	45.0	FOS/inulin
PFOS3	50.8		5.7	50.8	25.0	Without lactose
With galactooligosaccharides (PGOS)						
PGOS1	50.1	46.0	3.5	2.5[b]	32.0	GOS
PGOS2	55.0	–	1.5	1.5	51.0	GOS.mdx[c,d]
PGOS3	62.9	–	2.2	–	25.0	GOS.mdx
PGOS4	56.4	–	1.9	–	30.0	GOS.mdx
PGOS5	61.7	–	2.6	–	26.0	GOS.mdx
PGOS6	55.6	35.6	2.5	19.4	24.4	GOS.mdx
PGOS7	50.9	45.3	2.8	5.0	27.2	GOS.mdx
With galactooligosaccharides and fructooligosaccharides (PGOS/FOS)						
PGOS/FOS1	59.0	56.5	3.9		25.0	GOS/FOS
PGOS/FOS2	59.3	41.9	3.8	–	24.7	GOS/FOS.mdx[d]
PGOS/FOS3	59.3	41.9	3.8	–	24.7	GOS/FOS.mdx
PGOS/FOS4	59.3	41.9	3.8	–	24.7	GOS/FOS.mdx
PGOS/FOS5	55.7	19.7	3.8	–	23.0	GOS/FOS.mdx
PGOS/FOS6	53.6	–	2.9	–	–	GOS/FOS.mdx

[a] The data did not appear on the label.
[b] High molecular mass dextrin.
[c] mdx: Maltodextrins.
[d] Infant formula containing GOS.mdx and GOS/FOS.mdx were treated with α-amyloglucosidase.

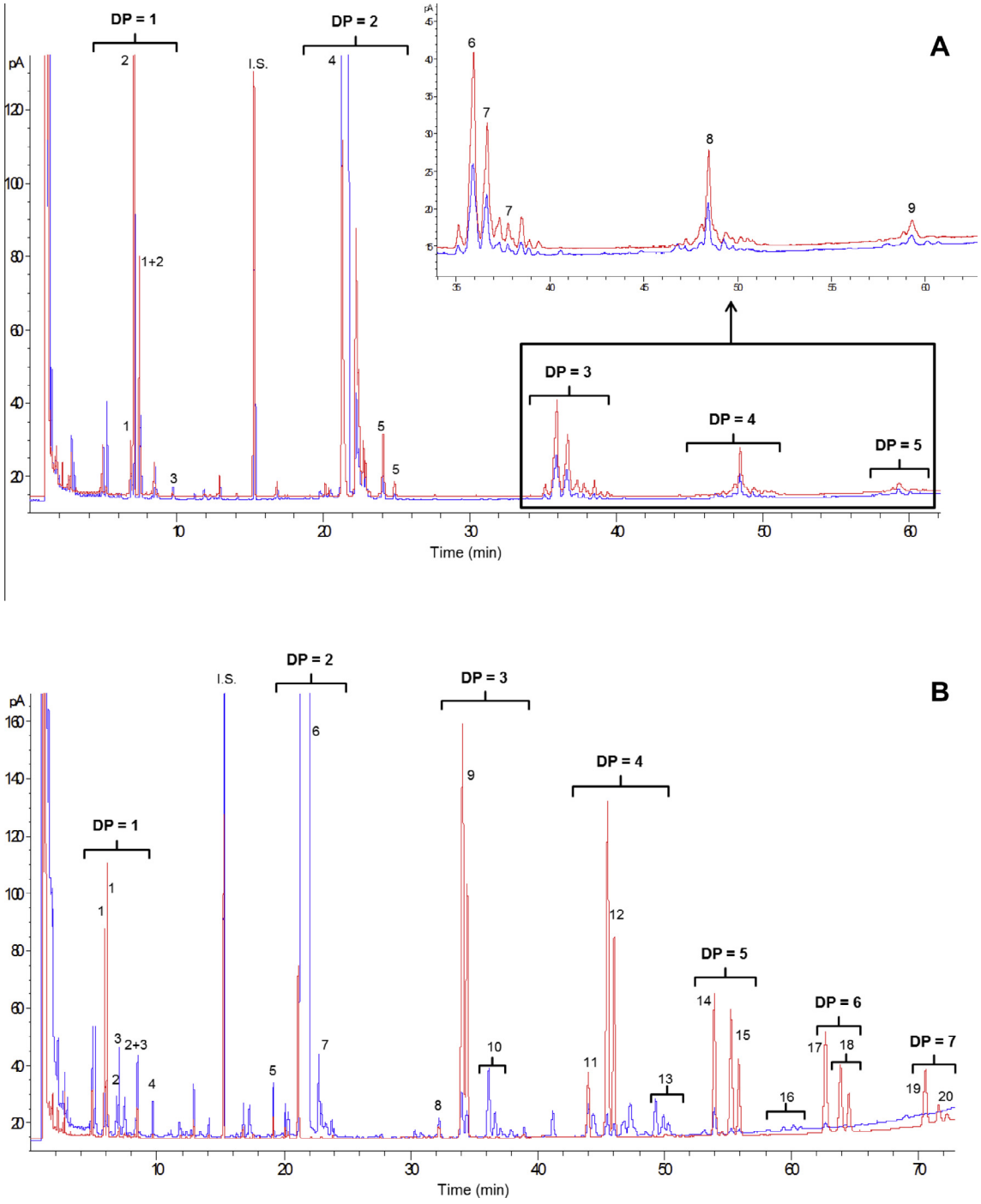


Fig. 1. GC-FID profiles of trimethylsilyl oximes (TMSO) of the carbohydrates present in (A): commercial Vivinal® GOS (red) and infant formula with GOS (blue). Peaks: 1: Galactose, 2: Glucose, 3: *Myo*-inositol, I.S.: Internal standard (phenyl-β-D-glucoside), 4: Lactose, 5: 6-Galactobiose, 6: 4′-Galactosyl-lactose, 7: 6′-Galactosyl-lactose, 8: 4′-Digalactosyl-lactose, 9: 4′-Trigalactosyl-lactose; and (B) commercial Raftilose® FOS (red) and infant formula with FOS (blue). Peaks: 1: Fructose, 2: Galactose, 3: Glucose, 4: *Myo*-inositol, 5: Sucrose, 6: Lactose, 7: Maltose, 8: Kestose, 9: Trifructosaccharides, 10: Maltotriose, 11: Nystose, 12: Tetrafructosaccharides, 13: Maltotetraose, 14: Fructosyl-nystose, 15: Pentafructosaccharides, 16: Maltopentaose, 17: Difructosyl-nystose, 18: Hexafructosaccharides, 19: Trifructosyl-nystose, 20: Heptafructosaccharides. DP: Degree of polymerisation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Wicklow, Ireland) (330 U/mL IF reconstituted at 100 mg/mL) in order to remove them.

3. Results and discussion

The GC method of Montilla et al. (2006) has been used in this study to determine carbohydrate composition of commercial IFs. The repeatability was calculated using an IF containing prebiotics and four replicates of this formula were analysed daily during the following 4 days (Table 2). As it can be observed, the repeatability was acceptable for all measured carbohydrates showing relative standard deviations (RSD) below 10%, with the exception of hexa and heptasaccharides (12.3%); these results were similar to those obtained by Montilla et al. (2006).

GC profiles of the carbohydrates found in the starting and follow-up IFs with prebiotics are depicted in Fig. 1. The used GC

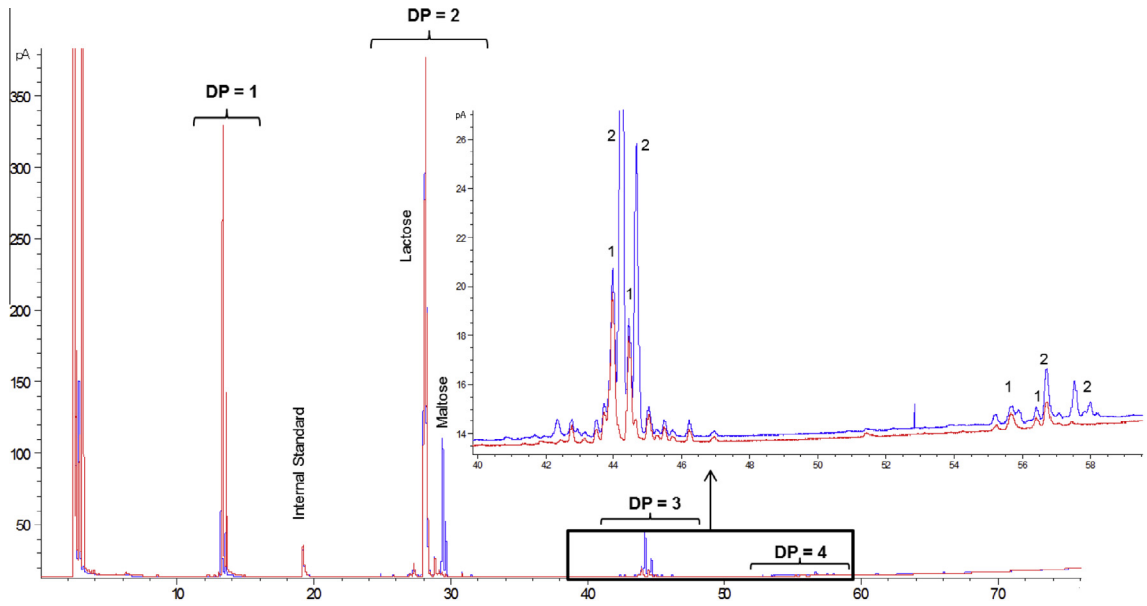


Fig. 2. GC-FID profiles of TMSO derivatives of carbohydrates present in a commercial infant formulas with maltodextrins and GOS before (blue) and after (red) of α-amyloglucosidase treatment. (1) GOS; (2) maltodextrins. DP: Degree of polymerisation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Carbohydrate content encountered (g/100 g) in the commercial infant formula under study (data shown as average value ± SD).

Product code	Fructose	Galactose	Glucose	*Myo*-inositol	Sucrose	Lactose	Maltose	GOS	FOS	Maltodextrins (DP 3–19)[a]	TCH[b]
C1		0.21 ± 0.00	0.27 ± 0.01	0.07 ± 0.00		70.1 ± 1.2					70.77
C2	0.02 ± 0.00	0.25 ± 0.01	0.39 ± 0.02	0.03 ± 0.00		30.7 ± 1.7	1.43 ± 0.06			11.7 ± 2.7	44.62
C3	0.01 ± 0.00	0.26 ± 0.00	0.15 ± 0.00	0.05 ± 0.00		45.5 ± 0.3	0.20 ± 0.00			0.8 ± 0.2	46.96
C4	0.05 ± 0.00	0.05 ± 0.00	1.59 ± 0.04	0.08 ± 0.01		22.2 ± 0.6	4.98 ± 0.11			22.5 ± 0.2	51.53
LF1	0.05 ± 0.00		1.27 ± 0.04	0.01 ± 0.00			5.73 ± 0.20			41.0 ± 0.6	48.14
LF2	0.04 ± 0.00	0.07 ± 0.00	0.99 ± 0.00	0.01 ± 0.00			3.01 ± 0.01			27.8 ± 5.7	31.88
LF3	0.03 ± 0.00		3.96 ± 0.04	0.03 ± 0.00			7.02 ± 0.05			45.7 ± 2.1	56.84
LF4	0.03 ± 0.00		1.53 ± 0.01	0.03 ± 0.00			5.11 ± 0.01			30.1 ± 1.6	36.85
PFOS1	0.34 ± 0.17	0.02 ± 0.00	0.19 ± 0.10	0.05 ± 0.00	0.17 ± 0.06	41.6 ± 1.7	0.30 ± 0.03		1.75 ± 0.07	4.3 ± 0.5	48.74
PFOS2	0.13 ± 0.01	0.08 ± 0.00	0.17 ± 0.01	0.06 ± 0.00	0.13 ± 0.01	47.5 ± 2.7	0.37 ± 0.00		1.57 ± 0.06	5.3 ± 0.7	55.27
PFOS3	0.05 ± 0.00	0.13 ± 0.01	0.69 ± 0.03	0.03 ± 0.00	0.19 ± 0.01		2.87 ± 0.13		5.00 ± 0.25	35.9 ± 0.2	44.90
PGOS1	0.03 ± 0.01	0.16 ± 0.01	1.29 ± 0.04	0.05 ± 0.00		44.5 ± 1.0	0.39 ± 0.01	3.16 ± 0.14			49.58
PGOS2	0.03 ± 0.00	0.23 ± 0.01	0.92 ± 0.02	0.06 ± 0.00		45.2 ± 0.9	0.32 ± 0.01	1.70 ± 0.04		1.7 ± 0.3	50.18
PGOS3	0.04 ± 0.00	0.13 ± 0.00	1.16 ± 0.00	0.04 ± 0.00		38.5 ± 0.0	1.57 ± 0.02	2.24 ± 0.05		11.3 ± 0.7	55.05
PGOS4	0.03 ± 0.00	0.12 ± 0.02	0.81 ± 0.09	0.04 ± 0.00		49.1 ± 0.0	0.54 ± 0.07	1.86 ± 0.14		2.3 ± 0.1	54.80
PGOS5	0.03 ± 0.00	0.15 ± 0.01	1.28 ± 0.13	0.04 ± 0.00		36.4 ± 3.5	1.57 ± 0.17	2.35 ± 0.05		11.1 ± 0.3	52.9
PGOS6	0.05 ± 0.00	0.13 ± 0.01	1.09 ± 0.09	0.02 ± 0.00		35.0 ± 3.1	0.82 ± 0.09	2.39 ± 0.07		14.5 ± 1.9	53.99
PGOS7	0.05 ± 0.00	0.16 ± 0.00	1.34 ± 0.02	0.05 ± 0.00		46.3 ± 3.2	0.72 ± 0.02	2.68 ± 0.14		4.2 ± 0.7	55.49
PGOS/FOS1	0.03 ± 0.00	0.12 ± 0.01	1.59 ± 0.14	0.03 ± 0.01	0.01 ± 0,00	51.1 ± 4.0		3.67 ± 0.11	0.13 ± 0.02		56.70
PGOS/FOS2	0.04 ± 0.00	0.20 ± 0.04	1.83 ± 0.16	0.04 ± 0.00	0.01 ± 0.00	39.9 ± 3.8	1.28 ± 0.24	3.73 ± 0.16	0.16 ± 0.00	26.2 ± 2.3	73.39
PGOS/FOS3	0.16 ± 0.21	0.17 ± 0.01	0.64 ± 0.14	0.03 ± 0.00	0.05 ± 0.05	38.5 ± 2.0	1.14 ± 0.12	3.64 ± 0.18	0.13 ± 0.01	15.0 ± 1.8	60.09
PGOS/FOS4	0.03 ± 0.00	0.12 ± 0.01	1.60 ± 0.16	0.03 ± 0.00	0.01 ± 0.00	33.9 ± 3.1	1.22 ± 0.09	3.79 ± 0.15	0.09 ± 0.00	11.1 ± 0.3	52.40
PGOS/FOS5	0.07 ± 0.01	0.18 ± 0.01	2.06 ± 0.10	0.03 ± 0.00	0.01 ± 0.00	19.2 ± 0.9	2.15 ± 0.11	3.70 ± 0.03	0.08 ± 0.01	19.4 ± 2.3	46.81
PGOS/FOS6	0.04 ± 0.00	0.11 ± 0.00	1.22 ± 0.02	0.07 ± 0.00	0.04 ± 0.00	29.5 ± 0.6	1.69 ± 0.03	2.28 ± 0.20	0.25 ± 0.05	21.5 ± 1.7	56.74

Product code: Like in Table 1.
[a] Values obtained by HPLC analysis.
[b] TCH: Total carbohydrates. Each value is the sum of fructose, galactose, glucose, *myo*-inositol, sucrose, lactose, maltose, GOS, FOS and maltodextrins (DP 3–19).

method allowed the quantification of fructose, galactose and glucose, *myo*-inositol, lactose, and prebiotic oligosaccharides GOS and FOS with a degree of polymerization (DP) of up to 7. Fig. 1 also shows carbohydrate profiles of commercial Vivinal® GOS (Fig. 1A) and Raftilose® FOS (Fig. 1B), which are commonly added to IF. It can be observed that the chromatographic profiles of commercial GOS and FOS were similar to those found in IFs, containing these carbohydrates. The chromatographic profiles of IFs containing GOS and maltodextrins after hydrolysis using α-amyloglucosidase are also shown in Fig. 2. In this case, the previous hydrolysis of maltodextrins allowed the quantification of GOS.

Table 3 shows the carbohydrate composition of the analysed commercial IFs. In general, fructose, glucose and galactose, were

detected in the majority of the studied IFs. Their concentrations were quite similar to those previously found in other IFs (Morales et al., 2004). Among them, glucose was the major monosaccharide being in highly variable amounts. Regarding to fructose, only three samples showed contents higher than 0.1 g/100 g, probably due to the presence of free fructose in the products with added FOS. The galactose content was higher than those found by Morales et al. (2004) and similar to those obtained by Troyano, Villamiel, Olano, Sanz, and Martinez-Castro (1996) in commercial sterilized milks. *Myo*-inositol, a polyalcohol that may play a significant role in the prevention of bronchopulmonary dysplasia and retinopathy in premature infants (Hallman, Saugstad, Porreco, Epstein, & Gluck, 1985), was present in all IFs at concentrations significantly higher than those reported for Spanish commercial milks (Troyano et al., 1996), but similar to the found by Woollard, Macfadzean, Indyk, McMahon, and Christiansen (2014). As expected, with the exception of the lactose-free IFs, lactose was the main carbohydrate in all studied products, with amounts varying between 19.2 and 70.1 g/100 g. Maltose was found in most of the analysed IFs in a range from 0.2 to 7 g/100 g. This compound is usually not added as ingredient in IFs and its presence may be related to the partial hydrolysis of the added maltodextrins during processing. Some low amounts of sucrose were found in the formulas enriched with FOS and in those enriched with GOS/FOS mixtures (0.13–0.19 and 0.01–0.05 g/100 g, respectively). Maltulose was also detected in five IFs in a range from 0.05 to 0.23 g/100 g (data not shown). This isomer of maltose was detected for the first time in IF by Morales et al. (2004), who found a wide variability from 0.13 to 0.8 g/100 g product, attributing this to differences in processing conditions.

The analysis of maltodextrins by HPLC-RID allowed the quantitative determination of oligosaccharides with DP of up to 19. As it can be seen from the Table 3, maltodextrins were the second major component (after lactose) of most of the studied products with contents varying from 0.84 to 45.7 g/100 g of product. Among them, the most abundant compounds were those with DP between 3 and 6 (data not shown).

In the IF containing prebiotic oligosaccharides the content of FOS and GOS were in the ranges of 1.6–5.0 and 1.7–3.2 g/100 g, respectively. In products enriched with FOS and GOS mixtures, the content was in the range of 0.08–0.25 and 2.3–3.8 g/100 g, respectively.

Some of these values and especially those referred to the products enriched with FOS were below the officially declared values on the labels. These differences are probably due to the difficulties of quantification of oligosaccharides with a high DP (inulins), since it was specified on the labels that these formula contained these polysaccharides.

The analysed IFs presented prebiotic oligosaccharide contents in the range of 2.0–6.4 g/L (considering IF reconstitution of 12.5% w/v). In Europe, IFs are supplemented with prebiotic oligosaccharides according to Directive 2006/141/CE, which allows addition of these carbohydrates in amounts of up to 8 g/L. Different studies have shown that this concentration limit is strictly respected and that the most commonly assayed prebiotics used in IFs are FOS, GOS and GOS/FOS mixtures (in a 9:1 ratio) at concentrations ranging between 1.5 and 8 g/L (Braegger et al., 2011). Several experimental studies have revealed that at these concentrations, prebiotics had a significant bifidogenic effect (Ben et al., 2008; Euler, Mitchell, Kline, & Pickering, 2005), modulated the intestinal flora and the immune system (Fanaro et al., 2005) and provided beneficial effects for formula-fed infants (Boehm et al., 2005). Also, the amounts of prebiotics added to IF were generally well tolerated and they did not produce adverse side effects (crying, regurgitation or vomiting) (Boehm et al., 2002; Closa-Monasterolo et al., 2013).

4. Conclusion

This study shows that the analysed IFs present carbohydrates content within the range indicated on the package labels. The combined utilisation of GC and HPLC techniques allows an overall quantification of the carbohydrate fraction, including prebiotic oligosaccharides. GC-FID and HPLC-RID are rapid, simple, cheap, and powerful analytical techniques commonly found in academic and industrial laboratories. In addition, GC presents high resolving power, sensitivity and selectivity which enables the determination of higher oligosaccharides in foods that are often present at low concentrations.

The results have demonstrated the usefulness of GC as a powerful tool for the analysis of mono-, and disaccharides, as well as GOS, FOS and their mixtures in IFs. Furthermore, the methodology proposed here may be used for routine quality control of IF and other food ingredients containing prebiotics.

Acknowledgments

This work has been supported by projects AGL2011-27884 and AGL2014-58205-REDC from Ministerio de Economía y Competitividad; ALIBIRD-CM S-2013/ABI-272 (Comunidad de Madrid).

References

Arslanoglu, S., Moro, G. E., Schmitt, J., Tandoi, L., Rizzardi, S., & Boehm, G. (2008). Early dietary intervention with a mixture of prebiotic oligosaccharides reduces the incidence of allergic manifestations and infections during the first two years of life. *Journal of Nutrition, 138*(6), 1091–1095.

Barile, D., & Rastall, R. A. (2013). Human milk and related oligosaccharides as prebiotics. *Current Opinion in Biotechnology, 24*(2), 214–219.

Ben, X. M., Li, J., Feng, Z. T., Shi, S. Y., Lu, Y. D., Chen, R., et al. (2008). Low level of galacto-oligosaccharide in infant formula stimulates growth of intestinal Bifidobacteria and Lactobacilli. *World Journal of Gastroenterology, 14*(42), 6564–6568.

Bode, L. (2009). Human milk oligosaccharides: Prebiotics and beyond. *Nutrition Reviews, 67*(s2), S183–S191.

Boehm, G., Lidestri, M., Casetta, P., Jelinek, J., Negretti, F., Stahl, B., et al. (2002). Supplementation of a bovine milk formula with an oligosaccharide mixture increases counts of faecal bifidobacteria in preterm infants. *Archives of Disease in Childhood. Fetal and Neonatal Edition, 86*(3), F178–F181.

Boehm, G., Stahl, B., Jelinek, J., Knol, J., Miniello, V., & Moro, G. E. (2005). Prebiotic carbohydrates in human milk and formulas. *Acta Paediatrica, 94*(s449), 18–21.

Braegger, C., Chmielewska, A., Decsi, T., Kolacek, S., Mihatsch, W., Moreno, L., et al. (2011). Supplementation of infant formula with probiotics and/or prebiotics: A systematic review and comment by the ESPGHAN committee on nutrition. *Journal of Pediatric Gastroenterology and Nutrition, 52*(2), 238–250.

Cilla, A., Lacomba, R., Garcia-Llatas, G., & Alegria, A. (2012). Prebiotics and nucleotides in infant nutrition; review of the evidence. *Nutrición Hospitalaria, 27*(4), 1037–1048.

Closa-Monasterolo, R., Gispert-Llaurado, M., Luque, V., Ferre, N., Rubio-Torrents, C., Zaragoza-Jordana, M., et al. (2013). Safety and efficacy of inulin and oligofructose supplementation in infant formula: Results from a randomized clinical trial. *Clinical Nutrition, 32*(6), 918–927.

Corzo-Martínez, M., Copoví, P., Olano, A., Moreno, F. J., & Montilla, A. (2013). Synthesis of prebiotic carbohydrates derived from cheese whey permeate by a combined process of isomerisation and transgalactosylation. *Journal of the Science of Food and Agriculture, 93*, 1591–1597.

Engfer, M. B., Stahl, B., Finke, B., Sawatzki, G., & Daniel, H. (2000). Human milk oligosaccharides are resistant to enzymatic hydrolysis in the upper gastrointestinal tract. *American Journal of Clinical Nutrition, 71*(6), 1589–1596.

Euler, A. R., Mitchell, D. K., Kline, R., & Pickering, L. K. (2005). Prebiotic effect of fructo-oligosaccharide supplemented term infant formula at two concentrations compared with unsupplemented formula and human milk. *Journal of Pediatric Gastroenterology and Nutrition, 40*(2), 157–164.

Fanaro, S., Boehm, G., Garssen, J., Knol, J., Mosca, F., Stahl, B., et al. (2005). Galacto-oligosaccharides and long-chain fructo-oligosaccharides as prebiotics in infant formulas: A review. *Acta Paediatrica, 94*(s449), 22–26.

Gibson, G. R., & Roberfroid, M. B. (1995). Dietary modulation of the human colonic microbiota: Introducing the concept of prebiotics. *Journal of Nutrition, 125*(6), 1401–1412.

Hallman, M., Saugstad, O. D., Porreco, R. P., Epstein, B. L., & Gluck, L. (1985). Role of myoinositol in regulation of surfactant phospholipids in the newborn. *Early Human Development, 10*(3–4), 245–254.

Knol, J., Scholtens, P., Kafka, C., Steenbakkers, J., Gross, S., Helm, K., et al. (2005). Colon microflora in infants fed formula with galacto- and fructo-

oligosaccharides: More like breast-fed infants. *Journal of Pediatric Gastroenterology and Nutrition, 40*(1), 36–42.
Montilla, A., van de Lagemaat, J., Olano, A., & del Castillo, M. D. (2006). Determination of oligosaccharides by conventional high-resolution gas chromatography. *Chromatographia, 63*(9–10), 453–458.
Morales, V., Olano, A., & Corzo, N. (2004). Ratio of maltose to maltulose and furosine as quality parameters for infant formula. *Journal of Agricultural and Food Chemistry, 52*(22), 6732–6736.
Moreno, F. J., Olano, A., Santa-Maria, C., & Corzo, N. (1999). Determination of maltodextrins in enteral formulations by three different chromatographic methods. *Chromatographia, 50*(11–12), 705–710.
Moro, G., Minoli, I., Mosca, M., Fanaro, S., Jelinek, J., Stahl, B., et al. (2002). Dosage-related bifidogenic effects of galacto- and fructooligosaccharides in formula-fed term infants. *Journal of Pediatric Gastroenterology and Nutrition, 34*(3), 291–295.
Roberfroid, M., Gibson, G. R., Hoyles, L., McCartney, A. L., Rastall, R., Rowland, I., et al. (2010). Prebiotic effects: metabolic and health benefits. *British Journal of Nutrition, 104*, S1–S63.
Troyano, E., Villamiel, M., Olano, A., Sanz, J., & Martinez-Castro, I. (1996). Monosaccharides and *myo*-inositol in commercial milks. *Journal of Agricultural and Food Chemistry, 44*(3), 815–817.
Urashima, T., Taufik, E., Fukuda, K., & Asakuma, S. (2013). Recent advances in studies on milk oligosaccharides of cows and other domestic farm animals. *Bioscience, Biotechnology, and Biochemistry, 77*(3), 455–466.
Woollard, D. C., Macfadzean, C., Indyk, H. E., McMahon, A., & Christiansen, S. (2014). Determination of *myo*-inositol in infant formulae and milk powders using capillary gas chromatography with flame ionisation detection. *International Dairy Journal, 37*(2), 74–81.

Research Article

Received: 17 January 2017 Revised: 6 April 2017 Accepted article published: 18 April 2017 Published online in Wiley Online Library: 9 June 2017

(wileyonlinelibrary.com) DOI 10.1002/jsfa.8384

An efficient process for obtaining prebiotic oligosaccharides derived from lactulose using isomerized and purified whey permeate

Carlos Sabater,[a] Agustín Olano,[a] Marin Prodanov,[b] Antonia Montilla[a]* and Nieves Corzo[a]

Abstract

BACKGROUND: One of the most promising uses of whey permeate (WP) is the synthesis of prebiotic oligosaccharides. Herein, commercial WP was submitted to chemical isomerization catalysed by sodium borate at an alkaline pH and subsequent purification using anion-exchange resins to remove boron. Subsequently, purified mixtures were used to synthesize prebiotic oligosaccharides using β-galactosidase from *Bacillus circulans*.

RESULTS: Isomerization of concentrated WP (200 g L^{-1} lactose) gave rise to levels of lactulose up to 155.5 g L^{-1} after 30 min of reaction (molar ratio of boron/lactose, 1/1; pH 12; 70 °C). Boron was removed from the isomerized WP (IWP) using the combination of a strong acid (IR-120, H^+) and a weak base (IRA-743) anion-exchange resins, reducing its level to <1 ppm, without loss of lactulose. During the transglycosylation reaction of purified IWP (lactose/lactulose ratio, 1/2.4) maximum content of prebiotic compounds was achieved, i.e. 690 g kg^{-1} WP after 3 h of reaction.

CONCLUSION: This study shows that combined chemical–enzymatic reactions together with the purification of IWP results in an efficient synthesis of prebiotic oligosaccharides.

Keywords: whey permeate; isomerization; anion-exchange purification; transglycosylation; galactooligosaccharides; oligosaccharides derived from lactulose

INTRODUCTION

Whey is the most abundant by-product generated from the cheese-making process. Approximately 80% of the volume of milk used to make cheese is recovered as whey, which contains more than half of the solids of milk. Traditionally, most of the whey produced has been treated as dairy wastewater in the industry, representing a major environmental problem (chemical oxygen demand of 50 000 mg L^{-1}, biological oxygen demand of 80 000 mg L^{-1}).[1,2] However, since the introduction of ultrafiltration techniques in the cheese industry to produce whey–protein concentrates, whey is more valued mainly for the nutritional benefits of the proteins recovered.[3,4] The remaining by-product, whey permeate (WP), is composed mainly of lactose (>70% on a dry matter basis), and it is of interest for application in different areas, mainly in the food and pharmaceutical industries; however, a considerable percentage of WP is still treated as dairy wastewater.

In recent years, one of the most promising uses of WP is the synthesis of prebiotic oligosaccharides such as galactooligosaccharides (GOS) via transglycosylation of lactose catalysed by β-galactosidases (EC 3.2.1.23) of different microbial origin.[5–8] Recent research has focused on enzymatic transglycosylation using lactulose or mixtures of lactose with other carbohydrates, giving rise to new bioactive oligosaccharides with different structures.[9] In these reactions, β-galactosidases use lactose as a galactosyl donor to transfer the galactosyl moiety to another lactose molecule or other suitable acceptor carbohydrate present in the mixture, giving rise to new β-linked galactosyl oligosaccharides.[10] However, there are no available data about oligosaccharide formation in transglycosylation of mixtures of lactose and lactulose standard.

Other prebiotic oligosaccharides obtained from WP are those derived from lactulose (OsLu), which have been synthesized by chemo-enzymatic methodologies by combination of chemical isomerization of lactose, using egg shell as catalyst, and subsequent transglycosylation of the resulting mixture using β-galactosidase from *Bacillus circulans*.[11] Moreover, OsLu have been produced by isomerization using sodium aluminates as catalyst, of transglycosylated mixtures previously obtained by hydrolysis and transglycosylation of lactose standard or from WP, using β-galactosidases from dairy *Kluyveromyces lactis* and *K. marxianus*.[12,13]

* *Correspondence to: A Montilla, Departamento de Bioactividad y Análisis de Alimentos, Instituto de Investigación en Ciencias de la Alimentación CIAL, (CSIC-UAM) CEI (UAM + CSIC), C/Nicolás Cabrera, 9, E-28049 Madrid, Spain. E-mail: a.montilla@csic.es*

a *Departamento de Bioactividad y Análisis de Alimentos, Instituto de Investigación en Ciencias de la Alimentación CIAL, Madrid, Spain*

b *Departamento de Producción y Caracterización de Nuevos Alimentos, Instituto de Investigación en Ciencias de la Alimentación CIAL, Madrid, Spain*

A wide number of catalysts, such as sodium and calcium hydroxide, complexation reagents (aluminates and borates), zeolites, sepiolites, egg and oyster shell,[11,14–18] have been used to isomerize lactose. The use of aluminates or borates enables yields up to 80% of lactulose.[19,20] However, the main drawback of the use of these two catalysts is that they must be removed before being used for human consumption. For many years, the World Health Organization (WHO) has recommended levels of boron lower than 0.5 mg L^{-1} in drinking water.[21]

Of the many separation techniques available for removal of borate, ion-exchange chromatography is one of the most widely used.[21] Amberlite IRA-743 resin is highly selective for borate anions and boric acid elimination. Boric acid reacts with 1-deoxy-1-methylamino-D-glucitol residues present in resin to yield a relatively stable borate complex. The combined use of this resin together with IR-120(H^+) has been used to remove boric acid (98%) from carbohydrate mixtures.[22]

The objective of this paper was to optimize an efficient methodology for improving a chemical–enzymatic synthesis of prebiotic mixtures derived from lactose (GOS and OsLu) safe for human consumption. The chemical–enzymatic process includes chemical isomerization of lactose from WP, using boron as a catalyst, and removal of boron by ion chromatography using selective ion-exchange resins. The purified isomerized WP (IWP) was further used to synthesize OsLu and GOS using β-galactosidase from *B. circulans*.

MATERIALS AND METHODS

Standards and reagents

Reagents employed for chromatographic analysis, including standards (D-glucose, D-galactose, D-fructose, D-lactose, lactulose, raffinose, stachyose, boric acid and β-phenylglucoside), were obtained from Sigma (St Louis, MO, USA). All other chemicals were of analytical grade. Amberlite IR-120(H^+) and Amberlite IRA-743 (free-base form) resins were purchased through Sigma. Ultrapure water (18.2 MΩ cm, with levels of 1–5 ng mL^{-1} total organic carbon and <0.001 EU mL^{-1} pyrogen) produced in-house with a laboratory water purification system (Milli-Q Synthesis A10, Millipore, Billerica, MA, USA) was used throughout.

β-Galactosidase from *B. circulans* (neutral lactase) was acquired from Biocon (Barcelona, Spain). β-Galactosidase activity was 245 U mL^{-1}, where 1 U is the amount of enzyme required to hydrolyse 1 µmol of *ortho*-nitrophenyl-β-galactoside (ONPG) per minute at a working temperature of 50 °C and pH 6.5 in 0.05 mmol L^{-1} phosphate buffer. Protein content determined using the Bradford method[23] was 92.1 g L^{-1}. Specific β-galactosidase activity was 2660 U g^{-1} of protein.

Physicochemical characterization of cheese WP

Industrial bovine WP powder with a lactose content of 810 g kg^{-1} was kindly supplied by the dairy company Reny Picot (Navia, Spain). The pH of reconstituted WP was measured using a pH meter (MP 230, Mettler-Toledo, Barcelona, Spain) at a concentration of 300 g kg^{-1}.

Isomerization of lactose from WP

Isomerization of WP was carried out following the method of Zokaee *et al.*[20] Solutions of WP containing 200 g L^{-1} lactose were prepared in water and boric acid (molar ratio boron/lactose, 1/1). Reaction was performed at 70 °C and pH was adjusted to 11 and 12 by adding 2 mol L^{-1} NaOH. Samples (30 mL) were then taken at 0, 10, 20, 30, 60, 90, 120, 150 and 180 min. The reaction was stopped by cooling in an ice-water bath and then centrifuged at 27 000 × *g* for 15 min at 4 °C. The supernatant was collected and stored at −18 °C until purification and further analysis. The isomerization reaction was performed in duplicate and each sample was analysed twice. Similar reactions were carried out using 2 mol L^{-1} $Ca(OH)_2$ at pH 12. Similarly, aliquots of 30 mL were taken at 0, 15, 30 and 45 min.

Boron removal

Boron was removed from the IWP using the combination of a strong acid (IR-120)(H^+) and a weak base (IRA-743) anion-exchange resins.[22] Weak base boron-selective resin (IRA-743) was regenerated by the following method: 14 mL of 1.2 mol L^{-1} HCl solution were slowly percolated through a glass column containing a bed volume of 7 mL IRA-743. The resin was then washed with 2 bed volumes of water, followed by 3.2 bed volumes of 1.1 mol L^{-1} NaOH and then washed with water until neutral pH. Strong acid anion-exchange resin IR-120(H^+) was regenerated using 2 bed volumes of 1.2 mol L^{-1} HCl solution percolated through a glass column containing a bed volume of 7 mL of IR-120(H^+) and washed with water until neutral pH. Columns with bed volumes of 63 mL of each resin were regenerated similarly.

Boron removal was confirmed by visible spectrophotometry (Azomethine H method)[24] and inductively coupled plasma mass spectrometry (ICP-MS), as indicated below.

Model systems

First, boron removal was studied in two different model systems containing 60 and 120 g L^{-1} total solids composed of boric acid (18% w/v), galactose (11% w/v), lactose (11% w/v) and lactulose (60% w/v), simulating IWP composition (Table 1). The pH of each solution was adjusted to pH 12 with 2 mol L^{-1} NaOH. An aliquot of 2 mL of each test solution was passed through a IR-120 (H^+) column (resin volume 7 mL) to remove sodium ions. Samples were eluted through the column with distilled water and collected in a final volume of 4 mL. These fractions were then passed through a IRA-743 column (resin volume 7 mL), eluted with distilled water and collected in a final volume of 11 mL. Flow rates were adjusted to 2 mL min^{-1}.

Isomerized cheese whey permeate

IWP was purified following the procedure described above using the same resins. Aliquots of 2 mL IWP solutions (237 g L^{-1} carbohydrates plus boric acid) and IWP solutions diluted twice (119 g L^{-1} carbohydrates plus boric acid) were used. Finally, two samples (18 mL) of IWP (isomerized for 20 and 30 min, 237 g L^{-1} carbohydrates plus boric acid) were passed through columns with IR-120(H^+) and IRA-743 resins (63 mL each). These samples were eluted with distilled water and lyophilized for subsequent use in transglycosylation reactions.

Synthesis of oligosaccharides derived from lactose (GOS) and lactulose (OsLu)

First, with the objective to optimize transglycosylation reactions, different assays using model systems prepared with lactose and lactulose in 0.05 mol L^{-1} sodium phosphate and pH 6.5 were performed. Different concentrations of lactose and lactulose were

Table 1. Lactose, lactulose, galactose and boron contents found in model systems of 60 and 120 g L^{-1} total solids and IWP of 120 and 240 g L^{-1} total solids after successive treatment with IR-120 (H+) and IRA-743 resins

		Concentration (g L^{-1} total solids)	Lactose (%)	Lactulose (%)	Galactose (%)	Boron(%)
Model systems	Before treatment	60	8.23 ± 0.06	67.22 ± 0.07	10.97 ± 0.19	13.59 ± 0.06
	Treatment with IR-120		8.65 ± 0.04	71.64 ± 0.74	8.61 ± 0.35	11.11 ± 0.43
	Treatment with IR-120 + IRA-743		11.11 ± 0.65	78.57 ± 0.01	9.84 ± 0.74	– *
	Before treatment	120	6.34 ± 0.24	68.66 ± 0.56	10.77 ± 0.11	14.24 ± 0.70
	Treatment with IR-120		7.64 ± 0.14	72.88 ± 0.47	9.00 ± 0.21	10.49 ± 0.39
	Treatment with IR-120 + IRA-743		9.81 ± 0.48	79.48 ± 0.25	10.37 ± 0.32	– *
IWP	Before treatment	120	20.71 ± 2.38	65.50 ± 0.96	2.62 ± 0.03	11.17 ± 1.38
	Treatment with IR-120		24.89 ± 0.48	70.48 ± 0.71	1.86 ± 0.03	2.78 ± 1.22
	Treatment with IR-120 + IRA-743		25.58 ± 0.42	72.91 ± 0.16	0.85 ± 0.02	– *
	Before treatment	240	17.35 ± 2.38	66.86 ± 0.96	2.66 ± 0.03	13.13 ± 1.38
	Treatment with IR-120		23.93 ± 0.39	68.39 ± 0.08	2.15 ± 0.20	5.53 ± 0.27
	Treatment with IR-120 + IRA-743		26.03 ± 0.05	72.22 ± 0.37	1.20 ± 0.09	– *

* Boron content under LOQ (0.02 g L^{-1}) HPLC-SEC-RID and confirmed by visible spectrophotometry (LOQ 0.001 g L^{-1}).

tested (250/250, 250/500 and 250/125, respectively, expressed as g L^{-1}), leading to three different molar ratios of lactose/lactulose (1/1, 1/2 and 2/1, respectively). Hydrolysis and transglycosylation of these model systems were performed by incubation at 50 °C with β-galactosidase from *B. circulans* (0.7 and 1.4 U mL^{-1}) in 0.05 mol L^{-1} sodium phosphate buffer (pH 6.5).

Transglycosylation reactions to obtain GOS and OsLu were performed using WP (equivalent to 250 g L^{-1} lactose), purified IWP (500 g L^{-1} carbohydrates and lactose/lactulose ratio, 1/2.4) and mixtures of WP–purified IWP (500 g L^{-1} carbohydrates; lactose/lactulose ratio, 1/1) as substrates which were reconstituted in water and adjusted to pH 6.5. Enzymatic synthesis was performed in a final volume of 1.5 mL in microtubes incubated in an orbital shaker at 50 °C and 38 × *g*. Samples were withdrawn at 0, 1, 3, 5, 8 and 24 h and immediately immersed in boiling water for 5 min to inactivate the enzyme. Samples were stored at −18 °C for subsequent analysis. Enzymatic reactions using 0.7 U mL^{-1} β-galactosidase from *B. circulans* were carried out in duplicate and each sample was analysed twice.

Analytical techniques

Gas chromatography with flame ionization detection (GC-FID)

The carbohydrate composition of isomerization and transglycosylation mixtures was determined by GC-FID. In isomerized samples, where boron interfered in the formation of oximes, carbohydrates were analysed as trimethylsilylated (TMS) ethers. First, a supernatant aliquot equivalent to 5–6 mg sugars was added to 0.2 mL internal standard (IS) solution (0.5 g L^{-1} phenyl-β-glucoside). The mixture was dried at 38–40 °C in a rotary evaporator (Büchi Labortechnik AG, Flawil, Switzerland). Dried samples were treated with 100 μL *N,N*-dimethylformamide and incubated at 70 °C for 30 min. Then, 150 μL *N*-trimethylsilylimidazole was added to obtain TMS ethers; the reaction was completed in 30 min at 70 °C. Silylated carbohydrates were extracted with 0.3 mL hexane and 0.3 mL water. Aliquots of 1 μL of the organic phase containing silyl derivatives were injected into the column.

GC-FID analyses of carbohydrates from boron-free transglycosylation mixtures were performed using trimethylsilylated oximes (TMSO) following the method of Brobst and Lott.[25] Sugar oximes were formed by adding 250 μL hydroxylamine chloride (2.5%) in pyridine to dried samples and heating the mixture at 70 °C for 30 min, then silylating with hexamethyldisilazane (250 μL) and trifluoroacetic acid (25 μL), maintaining at 50 °C for 30 min. Reaction mixtures were centrifuged at 6720 × *g* for 2 min at room temperature. Supernatants were injected or stored at 4 °C prior to analysis.

TMS and TMSO of carbohydrates were determined by GC using an Agilent Technologies 7890A gas chromatograph (Wilmington, DE, USA) equipped with a commercial fused-silica capillary column SPB-17, bonded, crosslinked phase (50% diphenyl/50% dimethylsiloxane; 30 m × 0.25 mm i.d., 0.25 µm film thickness) (Supelco, Bellefonte, PA, USA). The oven temperature was initially 200 °C; it was increased at a rate of 4 °C min^{-1} to 230 °C, then at a rate of 1 °C min^{-1} to 250 °C, then at a rate of 2 °C min^{-1} to 295 °C and held at this temperature for 50 min. The injector and detector temperatures were set at 280 and 295 °C, respectively. Injections were carried out in split mode (1/20) using nitrogen at 1 mL min^{-1} as carrier gas. Data acquisition and integration were performed using Agilent ChemStation Rev. B.03.01 software. To calculate the response factors relative to the IS, solutions containing glucose, galactose, lactose and lactulose were prepared over the expected concentration range in samples. Raffinose and stachyose were also used as standards of tri- and tetrasaccharides. The identities of carbohydrates were confirmed by comparison with relative retention times (RRT) of standard samples. Response factors were calculated after the duplicate analysis of standard solutions over the expected concentration range in samples. The amount of different carbohydrates and the yield of GOS and OsLu were expressed as percent weight of the total carbohydrate content in the reaction mixtures. All analyses were carried out in duplicate and data were expressed as mean ± standard deviation (SD).

Size exclusion high-performance liquid chromatography (HPLC) with refractive index detection (HPLC-SEC-RID)

Mono- and disaccharides, as well as boron from purified model systems and IWP, were analysed by HPLC-SEC-RID in a thermostated (80 °C) Shodex SUGAR KS-801 column (300 mm × 8 mm, 5 µm particle size) in combination with a Shodex SUGAR KS-G guard column (50 mm × 6 mm, 7 µm particle size) (Akzo Nobel, Brewster, NY, USA) using water as mobile phase in isocratic mode at a flow rate of 0.5 mL min^{-1}. Before analysis, model systems, IWP and purified IWP were diluted with water (~20 g L^{-1} carbohydrates) and

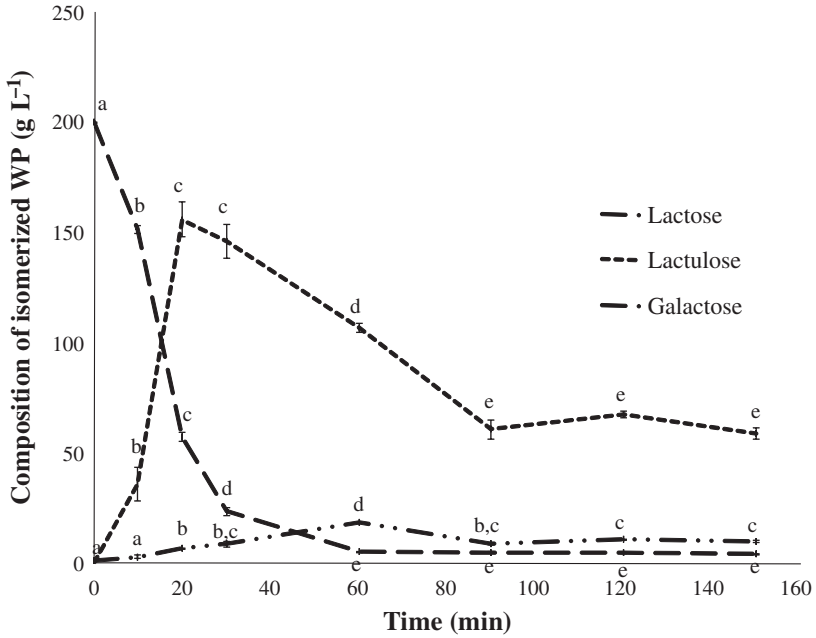


Figure 1. Evolution of lactose, lactulose and galactose (g L^{-1}) during the isomerization reaction using sodium borate as catalyst (lactose/boron molar ratio 1/1, pH 12, 70 °C). $^{a-e}$ Statistically significant differences considering individual carbohydrates ($P < 0.05$).

filtered using a 0.45 µm syringe filter (Symta). Analyses were performed in an Agilent Technologies 1220 Infinity LC System – 1260 RID (Böblingen, Germany). The injection volume was 50 μL (~1 mg total carbohydrates). Data acquisition and processing were performed using Agilent ChemStation software.

Boron and carbohydrates in the reaction mixtures were initially identified by comparing their retention times (RT) with those of standard compounds. Quantitative analysis was performed by the external standard method, using calibration curves in the range 0.02–2 g L^{-1} for lactose and lactulose, galactose and boron. All analyses were performed in duplicate, obtaining relative standard deviation (RSD) values below 10% in all cases. The amount of remaining carbohydrates and the yield of GOS and OsLu were expressed as percent by weight of the initial carbohydrate content in the reaction mixtures.

ICP-MS analysis

The total ion composition of the WP was determined using an ICP-MS ELAN 6000 PerkinElmer Sciex instrument at the Servicio Interdepartamental de Investigación (SIdI-UAM) of Madrid. Either a semi-quantitative or quantitative analysis of the elements of interest using the external calibration method and IS to correct instrumental drift was carried out. Values of major cations found in WP were: K, 30.7 mg g^{-1}; Na, 7.4 mg g^{-1}; Ca, 3.2 mg g^{-1}; and Mg, 1.4 mg g^{-1}. Boron content of IWP after purification was also determined using this methodology.

Analysis by matrix-assisted laser desorption/ionization time-of flight mass spectrometry (MALDI-TOF-MS)

Before MALDI-TOF-MS analysis, samples were submitted to a purification step following the method described by Morales *et al.*[26] Briefly, a total of 100 μL reaction mixture, containing 5 mg of carbohydrates, was dissolved in 10 mL of a 1% ethanol solution and stirred for 30 min with 300 mg activated charcoal Darco G60, 100 mesh (Sigma), to remove mono- and disaccharides. This mixture was vacuum filtered through Whatman No. 1 filter paper, and activated charcoal was washed with 50 mL of water. The oligosaccharides adsorbed onto the activated charcoal were extracted by stirring for 30 min with 10 mL ethanol–water solution (1/1, v/v) and then vacuum filtered. The ethanol–water solution was evaporated under vacuum at 40 °C. The sample was dissolved in 200 μL deionized water and filtered through 0.22 µm filters (Millipore Corp., Bedford, MA, USA) for further characterization by mass spectrometry.

The enriched fractions containing the oligosaccharides obtained by transglycosylation were characterized by MALDI-TOF-MS on a Voyager DE-PRO mass spectrometer (Applied Biosystems, Foster City, CA, USA) equipped with a pulsed nitrogen laser ($\lambda = 337$ nm, 3 ns pulse width and 3 Hz frequency) and a delayed extraction ion source. Ions generated by laser desorption were introduced into a TOF analyser (1.3 m flight path) with an acceleration voltage of 20 kV, 74% grid voltage, 0.001% ion guide wire voltage and delay time of 300 ns in the reflector positive ion mode. Mass spectra were obtained over the *m/z* range 100–1500.

Statistical analysis

ANOVA tests and Tukey's test for $P < 0.05$ (Statistix 8.1 software package) were applied to the carbohydrate content found in IWP mixtures in order to determine differences during the isomerization reaction.

RESULTS AND DISCUSSION

Isomerization of lactose from whey cheese permeate

Isomerization of WP was studied at pH values of 11 and 12. Yields near 40% were achieved at pH 11 after 240 min of isomerization, which was lower than the 70% obtained by Zokaee *et al.*[20] using aqueous lactose solutions at the same pH value; in the present study, WP containing a high number of cations was utilized. This could be attributed to an inhibitory effect of the high salt content in WP compared to buffer solutions. The influence of salts on

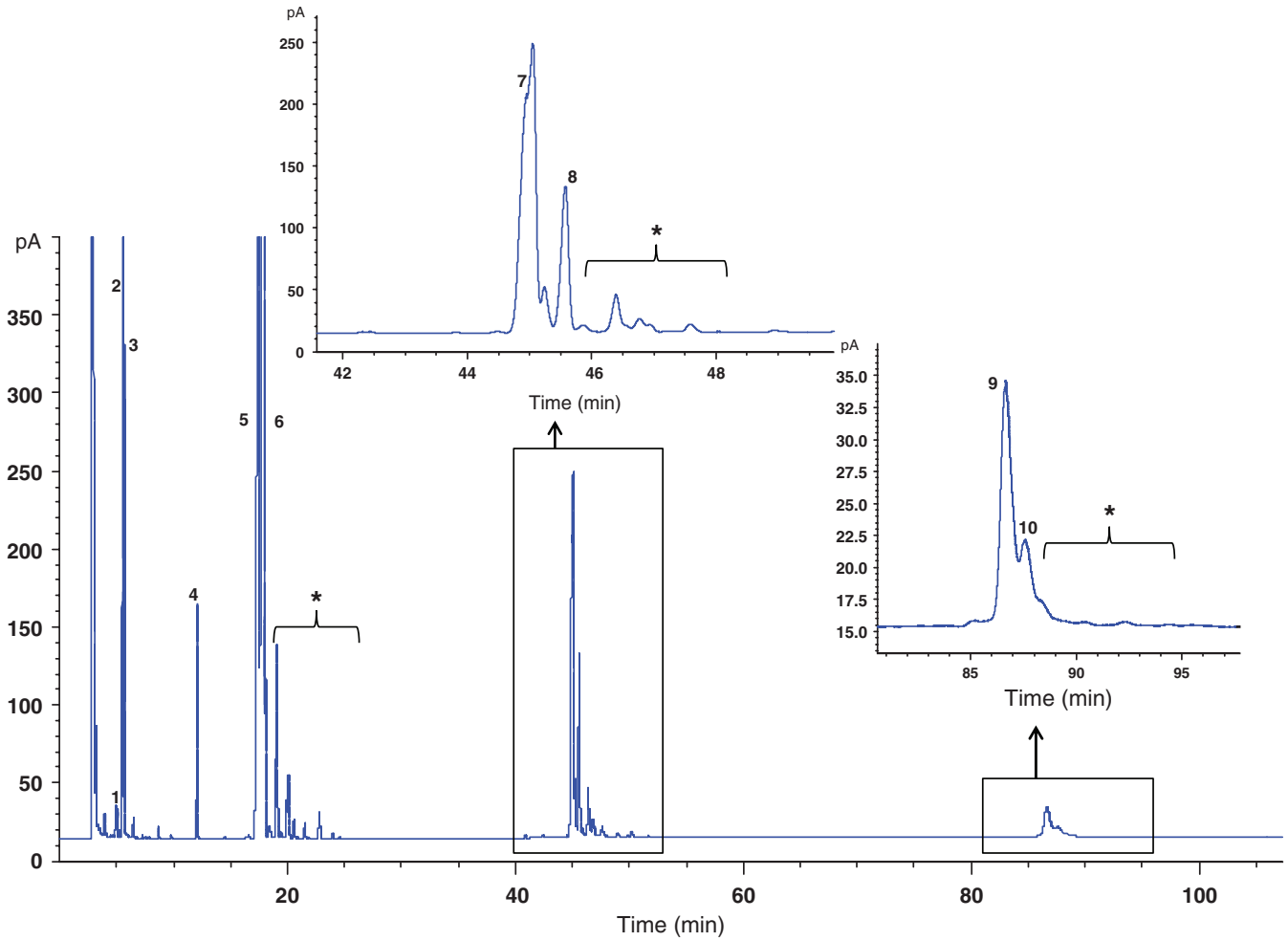


Figure 2. Chromatographic profile (GC-FID) of transgalactosylated lactose/lactulose model system using β-galactosidase from *Bacillus circulans* (pH 6.5, 50 °C, 3 h). Peaks: (1) fructose, (2) galactose, (3) glucose, (4) internal standard, (5) lactulose, (6) lactose, (7) 4′-galactosyl-lactulose, (8) 4′-galactosyl-lactose, (9) tentatively, 4′-digalactosyl-lactulose, (10) tentatively, 4′-digalactosyl-lactose; *unknown GOS.

lactose isomerization was previously observed by Montilla *et al.*[15] and Corzo-Martínez *et al.*[11]

When isomerization was carried out at pH 12, (Fig. 1) the optimal production of lactulose was 155.5 g L^{-1} (corresponding to 71% of total carbohydrates), and galactose and lactose contents were 6.4 and 57.1 g L^{-1}, respectively, after 20 min of reaction. No statistically significant differences ($P < 0.05$) were observed in galactose and lactulose content between 20 and 30 min of reaction. After 30 min, lactulose level rapidly decreased with the time of reaction, possibly due to its degradation to α- and β-isosaccharinic acids and galactose. Since no glucose or fructose was detected, it may be assumed that the reducing moiety of the disaccharide was first degraded into acidic compounds, followed by its hydrolysis, giving rise to galactose and isosaccharinic acids, which are further degraded into lower-molecular-weight organic acids.[15] The results obtained herein were similar to those obtained by Zokaee *et al.*[20]

In addition, another catalyst – calcium borate – was assayed to isomerize lactose to lactulose. Precipitation of calcium borate, which is less soluble than sodium borate, would enhance the removal of boron by centrifugation or filtration.[27] However, low yields of lactulose were obtained. The carbohydrate composition of the reaction mixture, under optimal conditions, was 13.9%, 64.7% and 21.4% of galactose, lactulose and lactose, respectively.

Ion-exchange removal of boron from IWP

To obtain a product safe for human consumption, efficient boron removal is necessary. For this purpose, IRA-743 resin has been used in combination with resin IR-120(H^+).[22] Model systems prepared at 60 and 120 g L^{-1} total solids (boric acid 14%, galactose 11%, lactose 8% and lactulose 67%, w/v,) and then adjusted to pH 12 with NaOH, were treated with 7 mL of each resin. According to the results obtained (Table 1), the concentration of solids was scarcely affected by the purification process. It can be observed that resin IR-120(H^+) produced no efficient boron retention, as expected, since this resin decreases the number of cations in the medium, exchanging H^+ for Na^+, neutralizing the reaction solution, and favouring subsequent boron elimination. Thus, when the two resins (IR-120, H^+ and IRA-743) were used, the remaining boron level was very low – below the limit of quantification (LOQ) – i.e. 0.02 g L^{-1}, and no important losses of lactulose were observed.

Considering these results, aliquots of IWP and IWP diluted twice (237 and 119 g L^{-1} carbohydrates and boric acid) were purified following the same treatment. The results obtained were similar to those found in model systems.

Finally, the purification process was scaled up. Undiluted IWP was passed through columns containing 63 mL of each resin and similar results were achieved, and remaining boron levels lower than

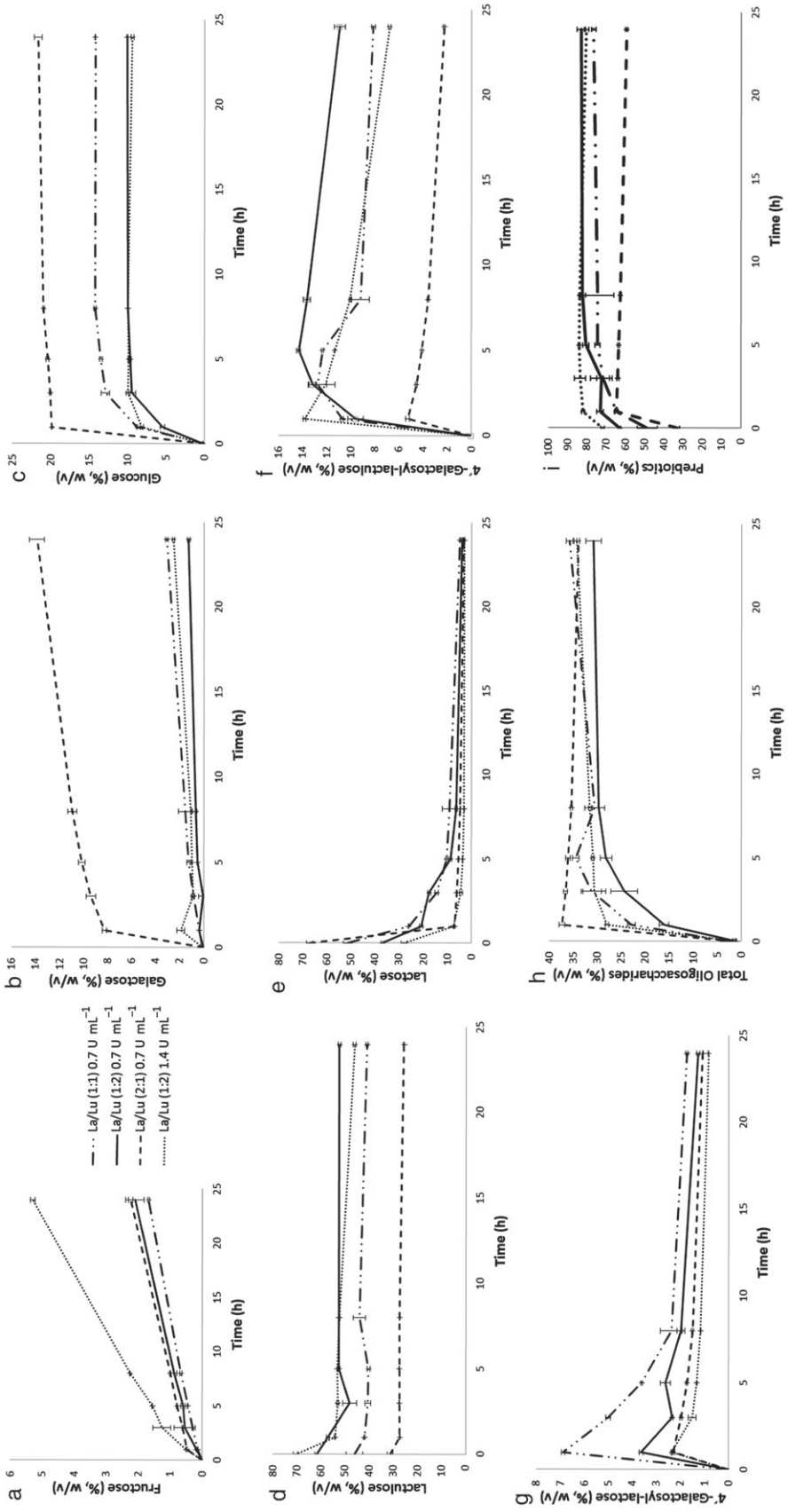


Figure 3. Evolution of fructose (a), galactose (b), glucose (c), lactulose (d), lactose (e), 4′-galactosyl-lactulose (f), 4′-galactosyl-lactose (g), GOS and OsLu (h) and total prebiotic carbohydrates (i) content during the enzymatic hydrolysis of lactose/lactulose (La/Lu) model systems with different molar ratios of lactose/lactulose (La/Lu) using 0.7 and 1.4 U mL^{-1} β-galactosidase from *Bacillus circulans*. GOS and OsLu include unknown di-GOS, 4′-galactosyl-lactulose, 4′-galactosyl-lactose and unknown tri- and tetra-GOS. Total prebiotic carbohydrates include GOS, OsLu and lactulose.

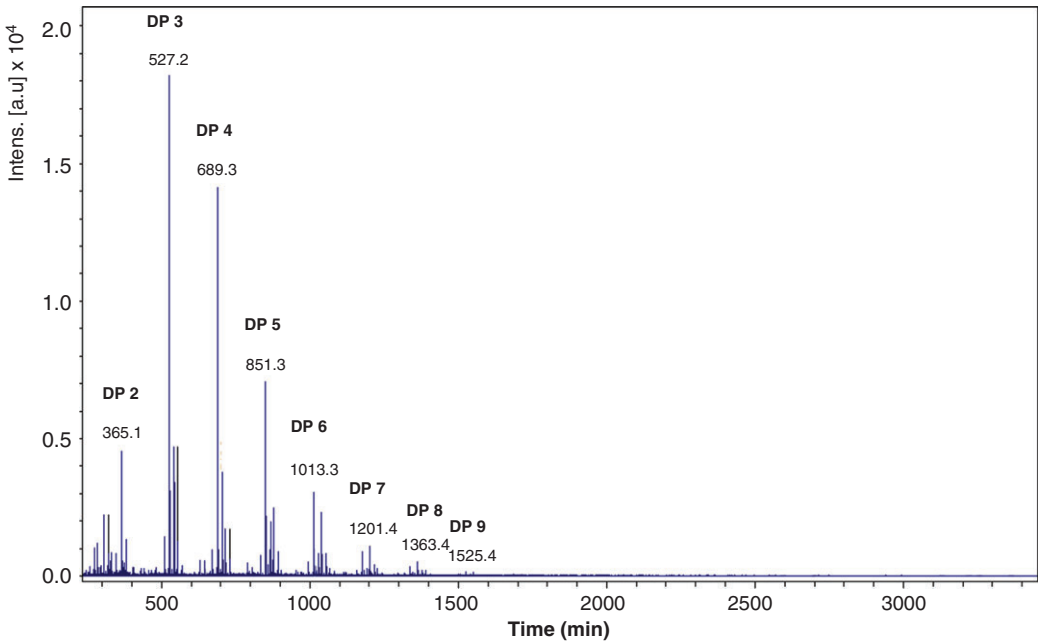


Figure 4. Mass spectra obtained by MALDI-TOF-MS of transgalactosylated IWP (lactose/lactulose, 1/1; pH 6.5, 50 °C, 3 h). DP, degree of polymerization.

1 ppm were found by visible spectrophotometry (Azomethine H method) and ICP-MS analysis. These results were better than to those reported by Hicks *et al.*,[22] who found a decrease of >99% of the original level in synthetic solutions of carbohydrates, with a final boron content of 216 ppm. Therefore, boron levels obtained in this study make possible the use of purified IWP as a food ingredient. Devirian and Volpe[28] reported that boron content of different fruits such as bananas and apples were between 21 and 43 ppm.

Transglycosylation of purified IWP using β-galactosidase from *Bacillus circulans*

First, reactions of transglycosylation were optimized using model systems of carbohydrates with different molar ratios of lactose/lactulose (such as has been described in the Materials and methods section). Figure 2 shows the chromatographic profile of carbohydrates found in reaction mixtures containing lactose and lactulose in a molar ratio of 1/1 after 3 h of reaction and using 0.7 U mL^{-1} β-galactosidase from *B. circulans*. The monosaccharides released – fructose, galactose and glucose (peaks 1, 2 and 3) by hydrolysis of lactose and lactulose – can be observed, as well as unreacted lactose and lactulose (peaks 5 and 6). These compounds were identified by comparison of their retention times with those of commercial standards. Moreover, the formation of GOS and OsLu (tri- and tetrasaccharides) with $\beta(1 \rightarrow 4)$ linkage was also detected since β-galactosidase from *B. circulans* leads to the formation of oligosaccharides with this linkage. Thus, 4′-galactosyl lactulose (peak 7) and 4′-galactosyl lactose (peak 8) were identified by comparison with the standards previously synthesized in our laboratory,[11] and 4′-digalactosyl lactulose and 4′-digalactosyl lactose (peaks 9 and 10) could only be tentatively identified. Unknown di-, tri- and tetrasaccharides were also detected (labelled with an asterisk in Fig. 2).

The quantification of oligosaccharides present in hydrolysis mixtures of system models containing lactose/lactulose ratios of 1/1, 1/2 and 2/1 is depicted in Fig. 3. During transglycosylation to produce GOS and OsLu, increasing amounts of glucose, galactose and fructose (Fig. 3 a–c) were released as a consequence of lactose and lactulose hydrolysis. Maximum release of glucose and galactose (21.6% and 13.9%, respectively) was reached with a lactose/lactulose ratio of 2/1, whereas maximum release of fructose was produced when the lactose/lactulose ratio was 1/2 and using 1.4 U mL^{-1} β-galactosidase (5.3%).

The different behaviour of disaccharides is evidenced by the amount of monosaccharides released; thus lactose is intensively hydrolysed (Fig. 3e) under all assayed conditions. After 24 h, it remained at <5%, whereas only about 10% of lactulose was hydrolysed. On the other hand, the minor amounts of galactose found, with respect to glucose, highlight that synthesized oligosaccharides may contain one or several galactose units.

With respect to the transglycosylation reaction, lactose and lactulose were very efficient acceptors, leading to the formation of GOS and OsLu. It could even be considered that lactulose is a more efficient acceptor than lactose, having in mind that the formation of 4′-galactosyl lactulose (14.3%) was greater than 4′-galactosyl lactose (6.9%) when both disaccharides were in the same proportion. Among the different lactose/lactulose ratios studied, i.e. 1/1, 1/2 and 2/1, the latter produced the highest amounts of prebiotic oligosaccharides (i.e. 37.3%, GOS and OsLu) after 1 h of reaction, using 0.7 U mL^{-1} β-galactosidase (Fig. 3h), mainly due to the large number of new disaccharides formed (20.1%). Under these conditions, the total prebiotic content, including lactulose, was 64.5% (Fig. 3i). Reaction mixtures containing a lactose/lactulose ratio of 1/2 produced 30.7% of prebiotics after 24 h of reaction and 0.7 U mL^{-1} β-galactosidase, and, considering lactulose, the level of prebiotics increased up to 82.9% after 24 h of reaction and using 0.7 and 1.4 U mL^{-1} enzyme. Finally, a lactose/lactulose ratio of 1/1 and using 0.7 U mL^{-1} β-galactosidase produced high quantities of prebiotics (76.4%). These data show the influence of type and concentration of carbohydrate present in initial reaction mixtures.

Subsequently, transglycosylation reactions using 0.7 U mL^{-1} β-galactosidase and non-isomerized WP and purified IWP at two lactose/lactulose ratios – 1/2.4 (normal ratio of two carbohydrates

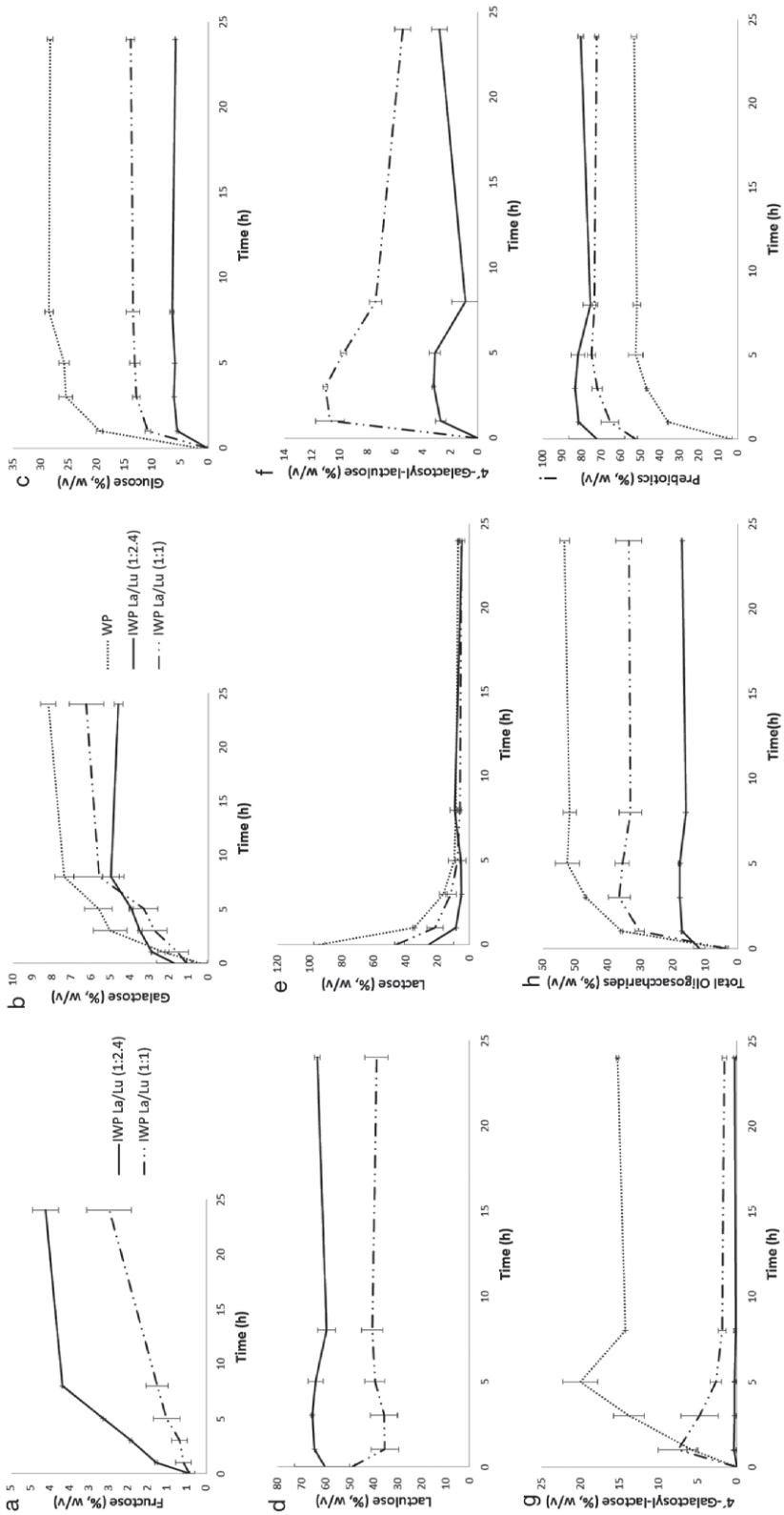


Figure 5. Evolution of fructose (a), galactose (b), glucose (c), lactulose (d), lactose (e), 4′-galactosyl-lactulose (f), 4′-galactosyl-lactose (g), GOS and OsLu (h) and total prebiotic carbohydrates (i) content during the enzymatic hydrolysis of WP and IWP with different molar ratios using 0.7 U mL^{-1} of *β*-galactosidase from *Bacillus circulans*. GOS and OsLu include unknown di-GOS, 4′-galactosyl-lactulose, 4′-galactosyl-lactose and unknown tri- and tetra-GOS. Total prebiotic carbohydrates include GOS, OsLu and lactulose.

after isomerization) and 1/1 – were carried out. Chromatographic profiles of oligosaccharides were similar to those obtained using standard solutions. MALDI-TOF-MS analysis of oligosaccharide-enriched fractions from non-isomerized WP and purified IWP transglycosylation mixtures enabled detection of oligosaccharides up to a degree of polymerization of 8 (*m/z* 1337) and 9 (*m/z* 1499), respectively (Fig. 4).

Evolution of carbohydrates from reaction mixtures originated by transglycosylation of WP, and IWP using β-galactosidase, is shown in Fig. 5. In general, the evolution of disaccharides (d, e) hydrolysis and the release of monosaccharides (a–c) were similar to what is observed in reactions carried out with lactose and lactulose standards. The level of 4′-galactosyl lactose found when WP was used as a substrate for transglycosylation was 20.1% of total carbohydrates, which is higher than that found in the case of standard solutions of lactose and lactulose or purified IWP (Fig. 5g). With respect to total oligosaccharides formed (GOS and OsLu), the WP sample showed the highest level (52.4% of total carbohydrates), after 5 h of reaction. IWP, on the other hand, with a lactose/lactulose ratio of 1/1, showed a higher content (36.4% total carbohydrates) than IWP (17.9%) after 1 h of reaction (Fig. 5h). However, considering prebiotic sugars (lactulose, GOS and OsLu), transglycosylation mixtures of IWP presented the highest amount (83.5% of total carbohydrates) after 3 h of reaction and a lactose/lactulose ratio of 1/2.4, and IWP with a lactose/lactulose ratio of 1/1 showed 75.0% after 5 h (Fig. 5i). These results agree with those reached in mixtures with lactose/lactulose ratios of 1/2 and 1/1. Finally, the highest yield of prebiotic carbohydrates was reached in IWP (1/2.4) with 690 g kg^{-1} permeate.

CONCLUSIONS

Most commercial prebiotic carbohydrates are obtained from enzymatic reactions. However, this study shows that combined chemical–enzymatic reactions, together with the purification of IWP, results in an efficient synthesis of prebiotic oligosaccharides. These combined reactions could be applied more broadly to the search for novel types of products using other organic by-products as substrates of chemical–enzymatic reactions in order to find new oligosaccharide structures with improved or complementary properties.

ACKNOWLEDGEMENTS

The authors acknowledge funding from the Spanish MINECO (AGL2014-53445-R), CSIC (COOPB-20099) and Spanish Danone Institute. Carlos Sabater acknowledges his FPU Predoc contract from Spanish MECD (FPU14/03619).

REFERENCE

1 Gänzle MG, Enzymatic synthesis of galacto-oligosaccharides and other lactose derivatives (hetero-oligosaccharides) from lactose. *Int Dairy J* **22**:116–122 (2012).
2 Seo YH, Park GW and Han JI, Efficient lactulose production from cheese whey using sodium carbonate. *Food Chem* **173**:1167–1171 (2015).
3 Reinbold R and Takemoto J, Use of Swiss cheese whey permeate by *Kluyveromyces fragilis* and mixed culture of *Rhodopseudomonas sphaeroides* and *Bacillus megaterium*. *J Dairy Sci* **71**:1799–1802 (1998).
4 Pródanov M, Tecnologías de separación por membranas para la revalorización de subproductos alimenticios, in *Retos medioambientales de la industria alimentaria.* Instituto Tomás Pascual para la Nutrición y la Salud, Madrid, pp. 103–124 (2012).
5 Cho YJ, Shin HJ and Bucke C, Purification and biochemical properties of a galactooligosaccharide producing β-galactosidase from *Bullera singularis*. *Biotechnol Lett* **25**:2107–2111 (2003).
6 Czermak P, Ebrahimi M, Grau K, Netz S, Sawatzki G and Pfromm PH, Membrane-assisted enzymatic production of galactosyl-oligosaccharides from lactose in a continuous process. *J Membrane Sci* **232**:85–91 (2004).
7 Goulas TK, Goulas AK, Tzortzis G and Gibson GR, Molecular cloning and comparative analysis of four β-galactosidase genes from *Bifidobacterium bifidum* NCIMB41171. *Appl Microbiol Biotechnol* **76**:1365–1372 (2007).
8 Lorenzen PC, Breiter J, Clawin-Rädecker I and Dau A, A novel bi-enzymatic system for lactose conversion. *Int J Food Sci Technol* **48**:1396–1403 (2013).
9 Díez-Municio M, Herrero M, Olano A and Moreno FJ, Synthesis of novel bioactive lactose-derived oligosaccharides by microbial glycoside hydrolases. *Microb Biotechnol* **7**:315–331 (2014).
10 Lu L, Xu X, Gu G, Jin L, Xiao M and Wang F, Synthesis of novel galactose containing chemicals by β-galactosidase from *Enterobacter cloacae* B5. *Bioresour Technol* **101**:6868–6872 (2010).
11 Corzo-Martínez M, Copoví P, Olano A, Moreno FJ and Montilla A, Synthesis of prebiotic carbohydrates derived from cheese whey permeate by a combined process of isomerisation and transgalactosylation. *J Sci Food Agric* **93**:1591–1597 (2013).
12 Cardelle-Cobas A, Corzo N, Villamiel M and Olano A, Isomerization of lactose-derived oligosaccharides: a case study using sodium aluminate. *J Agric Food Chem* **56**:10954–10959 (2008).
13 Padilla B, Frau F, Ruiz-Matute AI, Montilla A, Belloch C, Manzanares P *et al.*, Production of lactulose oligosaccharides by isomerisation of transgalactosylated cheese whey permeate obtained by β-galactosidases from dairy Kluyveromyces. *J Dairy Res* **82**:356–364 (2015).
14 Méndez A and Olano A, Lactulose: a review of some chemical properties and applications in infant nutrition and medicine. *Dairy Sci Abstr* **41**:531–535 (1979).
15 Montilla A, Del Castillo M, Sanz M and Olano A, Egg shell as catalyst of lactose isomerisation to lactulose. *Food Chem* **90**:883–890 (2005).
16 Aider M and de Halleux D, Isomerization of lactose and lactulose production: review. *Trends Food Sci Technol* **18**:356–364 (2007).
17 Olano A and Corzo N, Lactulose as a food ingredient. *J Sci Food Agric* **89**:1987–1990 (2009).
18 Seki N and Saito H, Lactose as a source for lactulose and other functional lactose derivatives. *Int Dairy J* **22**:110–115 (2012).
19 Hicks KB and Parrish FW, A new method for the preparation of lactulose from lactose. *Carbohydr Res* **82**:393–397 (1980).
20 Zokaee F, Kaghazchi T, Zare A and Soleimani M, Isomerization of lactose to lactulose: study and comparison of three catalytic systems. *Process Biochem* **37**:629–635 (2002).
21 Koseoglu P, Yoshizuka K, Nishihama S, Yuksel U and Kabay N, Removal of boron and arsenic from geothermal water in Kyushu Island, Japan, by using selective ion exchange resins. *Solvent Extr Ion Exch* **29**:440–457 (2011).
22 Hicks KB, Simpson GL and Bradbury AG, Removal of boric acid and related compounds from solutions of carbohydrates with a boron-selective resin (IRA-743). *Carbohydr Res* **147**:39–48 (1986).
23 Bradford MM, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**:248–254 (1976).
24 Evans S and Krähenbühl U, Boron analysis in biological material: microwave digestion procedure and determination by different methods. *Fresenius J Anal Chem* **349**:454–459 (1994).
25 Brobst K and Lott CJ, Determination of some components in corn syrup by gas–liquid chromatography of the trimethylsilyl derivatives. *Cereal Chem* **43**:35–43 (1966).
26 Morales V, Corzo N and Sanz M, HPAEC-PAD oligosaccharide analysis to detect adulterations of honey with sugar syrups. *Food Chem* **107**:922–928 (2008).
27 Irawan C, Kuo YL and Liu J, Treatment of boron-containing opto-electronic wastewater by precipitation process. *Desalination* **280**:146–151 (2011).
28 Devirian TA and Volpe SL, The physiological effects of dietary boron. *CRC Crit Rev Food Sci Nutr* **43**:219–231 (2003).

Journal of Food Composition and Analysis 66 (2018) 65–73

Contents lists available at ScienceDirect

Journal of Food Composition and Analysis

journal homepage: www.elsevier.com/locate/jfca

Original research article

Furosine and HMF determination in prebiotic-supplemented infant formula from Spanish market

C. Sabater, A. Montilla*, A. Ovejero, M. Prodanov, A. Olano, N. Corzo

Instituto de Investigación en Ciencias de la Alimentación CIAL, (CSIC-UAM) CEI (UAM + CSIC), C/Nicolás Cabrera, 9, E-28049 Madrid, Spain

ARTICLE INFO

Chemical compounds studied in this article:
Furosine (PubChem CID: 92043346)
5-Hydroxymethylfurfural (PubChem CID: 237332)

Keywords:
Furosine
HMF
Prebiotics
Infant formula
Storage
Food analysis
Support vector machines
Random forests

ABSTRACT

Evaluation of Maillard reaction (MR) in commercial prebiotic-supplemented and not supplemented infant formulas (IFs) was carried out through determination of furosine (2-furoylmethyl-ε-lysine) and hydroxymethylfurfural (HMF). Furosine was present in all studied IFs, ranging from 94 to 1226 and 315 to 965 mg $100\,g^{-1}$ of protein, in samples without prebiotics and prebiotic-supplemented IFs respectively. HMF was found in all IFs in the range 62–510 μg $100\,g^{-1}$ product. No statistical differences in HMF and furosine contents between prebiotic-supplemented and not supplemented IFs were observed. Storage of six representative IFs, with and without prebiotics, at room temperature for 15 months, did not produce changes in HMF content after 8 months of storage, while furosine content increased significantly throughout the storage time. The high amounts of furosine found in some IFs may be attributed to excessive heat treatment during processing, inadequate storage of IFs or ingredients used in their manufacturing process. The use of furosine and HMF as thermal indicators allows evaluation of the quality of prebiotic-supplemented IFs, which was similar to IFs without prebiotics. Finally, two supervised classification methods (support vector machines and random forests) were applied to classify IFs according to protein and carbohydrate source and IF type.

1. Introduction

The food industry has made several attempts to develop infant formulas (IFs) that fulfil the nutritional needs of newborn or young infants. This is particularly important in instances where mothers cannot produce enough milk. In recent years IFs have been supplemented with prebiotic oligosaccharides as growth-promoting factors for bifidobacteria, which are widely recognised as beneficial for newborn health (Ben et al., 2008; Knol et al., 2005; Fanaro et al., 2005). IFs composition includes as major components carbohydrates (54–61 g $100\,g^{-1}$ of product) and proteins (11–15 g $100\,g^{-1}$ of product). Sugars, such as corn syrups, lactose, sucrose or starch, are used to formulate IFs. They can also be supplemented with prebiotics, such as lactulose, galactooligosaccharides (GOS), fructooligosaccharides (FOS), mixtures GOS/FOS, polydextrose; isomaltooligosaccharides or xylooligosaccharides (Ackerman et al., 2017). Protein source can be varied, although the most used is whey protein (Morales et al., 2004).

The manufacture of IFs includes different stages, such as blending of ingredients, homogenisation, pasteurisation, spray-drying and storage that have a great influence in their final quality (Morales et al., 2004). Some of these processes, involving heat, reactions and/or interactions between constituents can give rise to a loss of nutritive value; this is very important because IFs sometimes are the only source of infant nutrition during the first months of life. Maillard reaction (MR) is one of the main reactions causing deterioration of proteins during processing or storage of foods. Since IFs contain high concentrations of carbohydrates and proteins, MR plays an important role during elaboration from the point of view of losses of nutritive value.

MR is a very complex reaction that covers transformations that produces a large number of the so-called MR products (Kim, 2010). This reaction takes place during processing of foods, between the carbonyl group of a reducing sugar and the free amino group of proteins, peptides or amino acids; therefore it plays an important role in IFs, due to their carbohydrate and protein composition, affecting adversely the protein quality by loss of the bioavailability of essential amino acids (Nasirpour et al., 2006). Loss of available lysine is the most negative nutritional consequence of the MR in IFs (Michalak et al., 2006). Storage of powdered IFs can also promote MR to different extents, depending on environmental conditions (temperature and humidity) and product composition (Michalak et al., 2006).

Different compounds have been selected as indicators of different stages of MR, furosine (2-furoylmethyl-ε-lysine, 2-FM-Lys) being one of the most used as an indicator of the early stages of MR in processed foods. Furosine determination is an indirect measure of Amadori

* Corresponding author.
E-mail address: a.montilla@csic.es (A. Montilla).

https://doi.org/10.1016/j.jfca.2017.12.004
Received 7 April 2017; Received in revised form 20 November 2017; Accepted 5 December 2017
Available online 06 December 2017
0889-1575/

Table 1
Dry matter (DM), water activity (a_w), pH and protein determined in the infant formula under study (n = 2, data shown as mean ± SD).

code	dry matter (%)	a_w	pH	% protein (N × 6.25)	protein source[a]	carbohydrates[b]
Infant formula without prebiotics						
conventional						
L1 (S)	95.4 ± 0.5	0.15	6.74	9.6 ± 0.2	whey protein	lactose
L2 (S)	99.2 ± 0.0	0.11	6.93	10.1 ± 0.0	whey protein	lactose, MDX[c]
L3 (F)	96.1 ± 0.0	0.12	6.83	12.1 ± 0.0	whey protein hydrolysate	lactose, starch
L4 (F)	98.3 ± 0.0	0.14	6.70	16.0 ± 0.1	milk protein	lactose, MDX
without lactose						
LF 1(S)	96.8 ± 0.0	0.13	6.67	10.4 ± 0.2	milk protein	MDX
LF 2(F)	97.9 ± 0.2	0.07	6.74	11.1 ± 0.1	milk protein	MDX
LF3 (S)	98.9 ± 0.0	0.13	7.12	12.1 ± 0.5	soy protein	MDX
LF4 (S)	97.3 ± 0.1	0.11	6.83	12.1 ± 0.3	soy protein	MDX
Infant formula with prebiotics						
with fructooligosaccharides						
PFOS1 (S)	96.8 ± 0.0	0.21	6.89	9.8 ± 0.0	whey protein	lactose, FOS[d]/inulin
PFOS2 (F)	96.6 ± 0.2	0.17	6.80	12.8 ± 0.0	whey protein	lactose, FOS/inulin
PFOS3 (S)	97.7 ± 0.4	0.11	6.96	11.6 ± 0.4	whey protein	without lactose, FOS
with galactooligosaccharides						
PGOS1 (S)	98.0 ± 0.1	0.10	7.04	10.6 ± 0.0	whey protein	lactose, GOS[e]
PGOS2 (F)	97.8 ± 0.2	0.16	7.22	10.4 ± 0.1	whey protein	lactose, GOS, MDX
PGOS3 (F)	96.5 ± 0.2	0.13	7.00	10.3 ± 0.5	whey protein	lactose, GOS, MDX
PGOS4 (S)	98.5 ± 0.4	0.12	6.92	10.5 ± 0.2	whey protein	lactose, GOS, MDX
PGOS5 (F)	94.5 ± 0.6	0.13	6.83	12.1 ± 0.2	whey protein	lactose, GOS, MDX
PGOS6 (S)	96.7 ± 0.4	0.15	7.13	10.7 ± 0.1	milk protein	lactose, GOS, MDX
PGOS7 (S)	96.6 ± 0.4	0.13	7.16	10.5 ± 0.0	whey protein, amino acids	lactose, GOS, MDX
with GOS and FOS (all samples were follow-up)						
PGOS/FOS1	96.5 ± 0.2	0.20	6.91	9.2 ± 0.0	whey protein hydrolysate	lactose, GOS/FOS
PGOS/FOS2	98.5 ± 0.1	0.14	7.12	9.6 ± 0.1	whey protein	lactose, GOS/FOS, MDX
PGOS/FOS3	95.0 ± 0.5	0.28	6.80	8.3 ± 0.3	whey protein	lactose, GOS/FOS, MDX
PGOS/FOS4	95.6 ± 0.8	0.24	6.68	9.3 ± 0.2	whey protein	lactose, GOS/FOS, MDX
PGOS/FOS5	98.4 ± 0.2	0.11	6.99	11.2 ± 0.3	amino acids	lactose, GOS/FOS, MDX
PGOS/FOS6	97.5 ± 0.8	0.14	7.23	12.3 ± 0.2	whey protein	lactose, GOS/FOS, MDX

S: starting IFs. F: follow-up IFs.
[a] Data on the label.
[b] Sabater et al. (2016).
[c] MDX, maltodextrins.
[d] FOS, fructooligosaccharides.
[e] GOS, galactooligosaccharides.

compounds, which are formed in the first stages of MR and its determination provides information about early quality changes caused in treated or stored foods (Boitz and Mayer, 2016; Corzo-Martínez et al., 2012; Fratianni et al., 2017; Gokmen et al., 2008; Ríos-Ríos et al., 2018).

Another indicator used to evaluate intermediate and advanced stages of MR in processed foods is hydroxymethylfurfural (HMF) (Albalá-Hurtado et al., 1997; Michalak et al., 2006). In acid media, HMF is also formed by degradation of hexoses and was one of the first compounds to be used as an indicator to evaluate the severity of heat treatment or length of storage in several foods, such as milk (Guerra-Hernández et al., 2002b), orange juice, honey, breakfast cereals, biscuits and jam (Teixido et al., 2006). Depending on the procedure chosen to measure HMF, it can be determined either as free HMF or free HMF plus potential HMF derived from other browning intermediates (total HMF) (Morales et al., 1996).

A lot of data are available about the use of the above mentioned indicators to assess the quality of liquid and powder IFs (Chávez-Servín et al., 2015; Contreras-Calderón et al., 2009; Ferrer et al., 2005; Guerra-Hernández et al., 2002a,b; Michalak et al., 2006; Roux et al., 2016). However, in IFs supplemented with prebiotics, the studies are mainly focused on the influence of these carbohydrates on growth and promotion of gut health of infants (Ackerman et al., 2017) or to know carbohydrate composition (Sabater et al., 2016). Few studies exist regarding the relationship between IFs and indicators (Contreras-Calderón et al., 2017), and the usefulness of quality parameters to classify samples.

Therefore, the aim of this study was to evaluate quality of IFs supplemented with prebiotics, compared to others without prebiotics, through thermal indices, furosine and free HMF determination, as well as the influence of storage of IFs (8 and 15 months) on MR. In addition, in order to study the relationships between the IFs studied grouped in different categories (starting and follow-up IFs, IFs containing proteins from different origin and different main carbohydrates) and quality parameters here determined, two supervised classification methods based on statistical learning theory have been applied: support vector machines (SVM) and random forests (RF). They are promising methods applied on spectroscopic data of milk powder for brand identification and component analysis (Wu et al., 2008), and MS data for characterisation and classification of cheese samples (Fabris et al., 2010).

2. Materials and methods

2.1. Chemicals

All chemicals used were of analytical grade. Hydrochloric acid, acetic acid and potassium chloride were obtained from Panreac (Barcelona, Spain). Sep-Pak cartridges (C_{18}) were purchased from Waters Millipore (Milford, MA). Ultrapure water quality (18.2 MΩ cm) with 1 − 5 ppb total organic carbon (TOC) and < 0.001 EU mL^{-1} pyrogen levels was produced in-house using a Milli-Q Synthesis A10 laboratory water purification system from Millipore (Billerica, MA).

2.2. Samples

Different types of commercial powdered IFs (n = 24), starting and follow-up formulas, were purchased from local markets. Sixteen IFs containing prebiotics i.e. fructooligosaccharides (PFOS, n = 3),

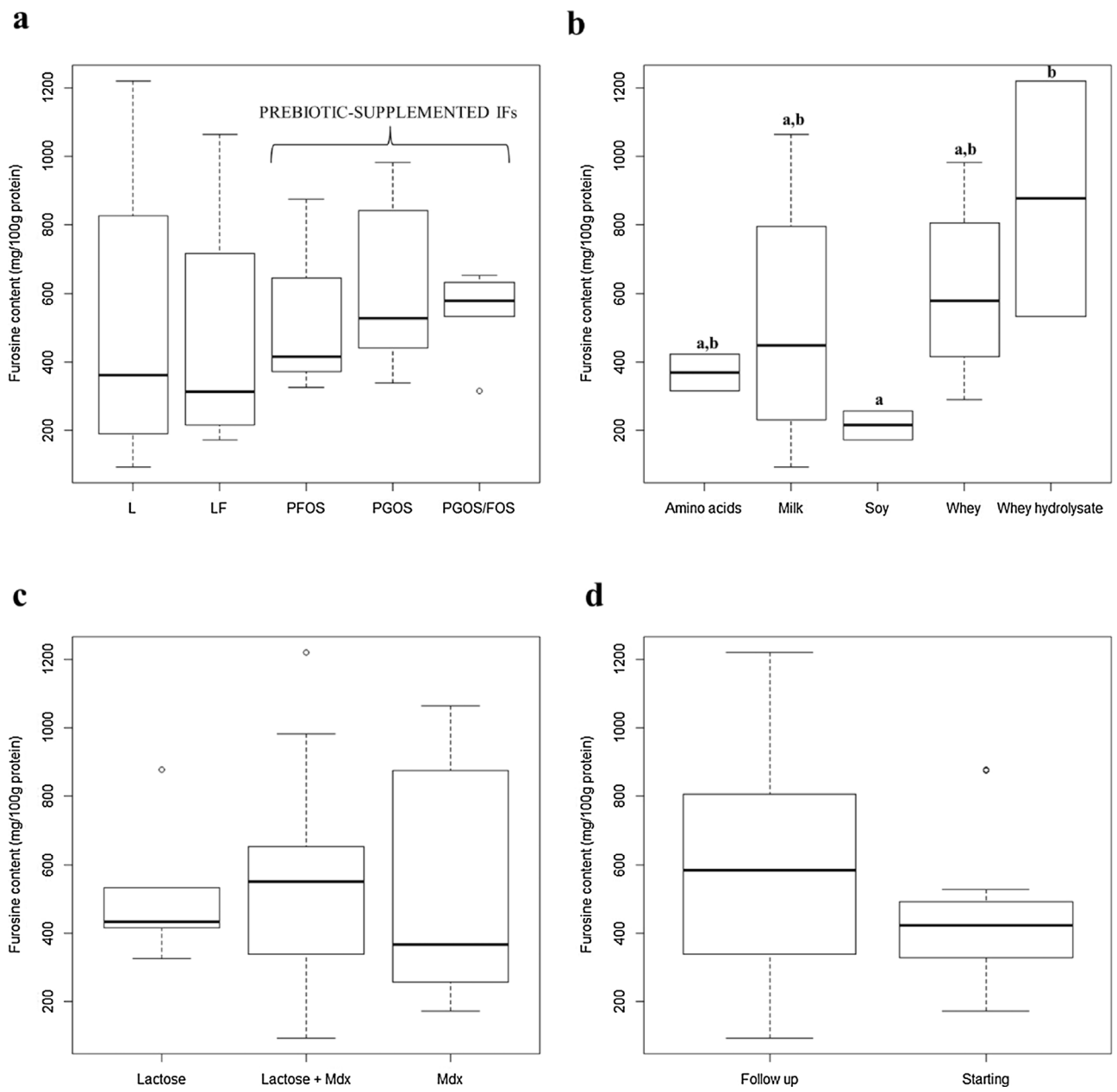


Fig. 1. Furosine content found in infant formulas (IFs): **(a)** with lactose (L, $n = 4$), lactose-free (LF, $n = 4$) and prebiotic-supplemented (with fructooligosaccharides, PFOS $n = 3$; with galactooligosaccharides, PGOS $n = 7$ and PGOS/FOS $n = 6$); **(b)** according to their protein source: amino acids ($n = 2$), milk ($n = 4$), soy ($n = 2$), whey ($n = 14$) and whey hydrolysate ($n = 2$); **(c)** according to their carbohydrate source: lactose ($n = 5$), maltodextrins (MdxMDX, $n = 5$), lactose + MdxMDX ($n = 14$); and **(d)** by age, follow-up ($n = 13$) and starting ($n = 11$). (Fig. 1b, ^{a,b} Statistically significant differences ($p < 0.05$) between IFs).

galactooligosaccharides (PGOS, $n = 7$), and mixtures of galactooligosaccharides and fructooligosaccharides (PGOS/FOS, $n = 6$), and lactose (except PFOS3 sample) were studied. The other eight IFs were without prebiotics; four of them contained lactose and maltodextrins or starch (L) while the other four were lactose-free (LF) but contained maltodextrins (Sabater et al., 2016). Before analysis, IFs were stored at 4 °C and analysed before the sell-by date. Also, six representative IFs were stored for 15 months at room temperature 25 °C (20–33 °C) and humidity between 17 and 43%.

Also, for statistical analysis IFs were categorised according to three criteria i) proteins source: whey ($n = 14$), milk ($n = 4$), hydrolysed whey ($n = 4$), and soy ($n = 2$); ii) considering carbohydrates data of these IFs previously published (Sabater et al., 2016), we have categorised samples with lactose as main carbohydrate ($n = 5$), samples with lactose and maltodextrins ($n = 14$) and with MdxMDX ($n = 5$); and iii) by age: starting ($n = 11$) and follow-up ($n = 13$) IFs.

2.3. Infant formula characterisation

All IFs were reconstituted in water at 10% (w/v) and the pH values were measured using a Mettler Toledo Five Easy Plus pH Meter. The dry matter (DM) content was determined gravimetrically by drying the samples until constant weight according to the AOAC method 950.01 (AOAC, 1990a). Water activity (a_w) measurement was carried out in an AW Sprint TH-500 instrument (Novasina, Pfäffikon, Switzerland). Saturated aqueous solutions of LiCl, $MgCl_2$, $Mg(NO_3)_2$, NaCl, $BaCl_2$ and $K_2Cr_2O_7$ were used to calibrate the sensor unit. The Kjeldahl method was used to determine total nitrogen (TN) using 6.25 as conversion factor (TN × 6.25), following the AOAC method 920.165 (AOAC, 1990b). All determinations were carried out in duplicate and data were

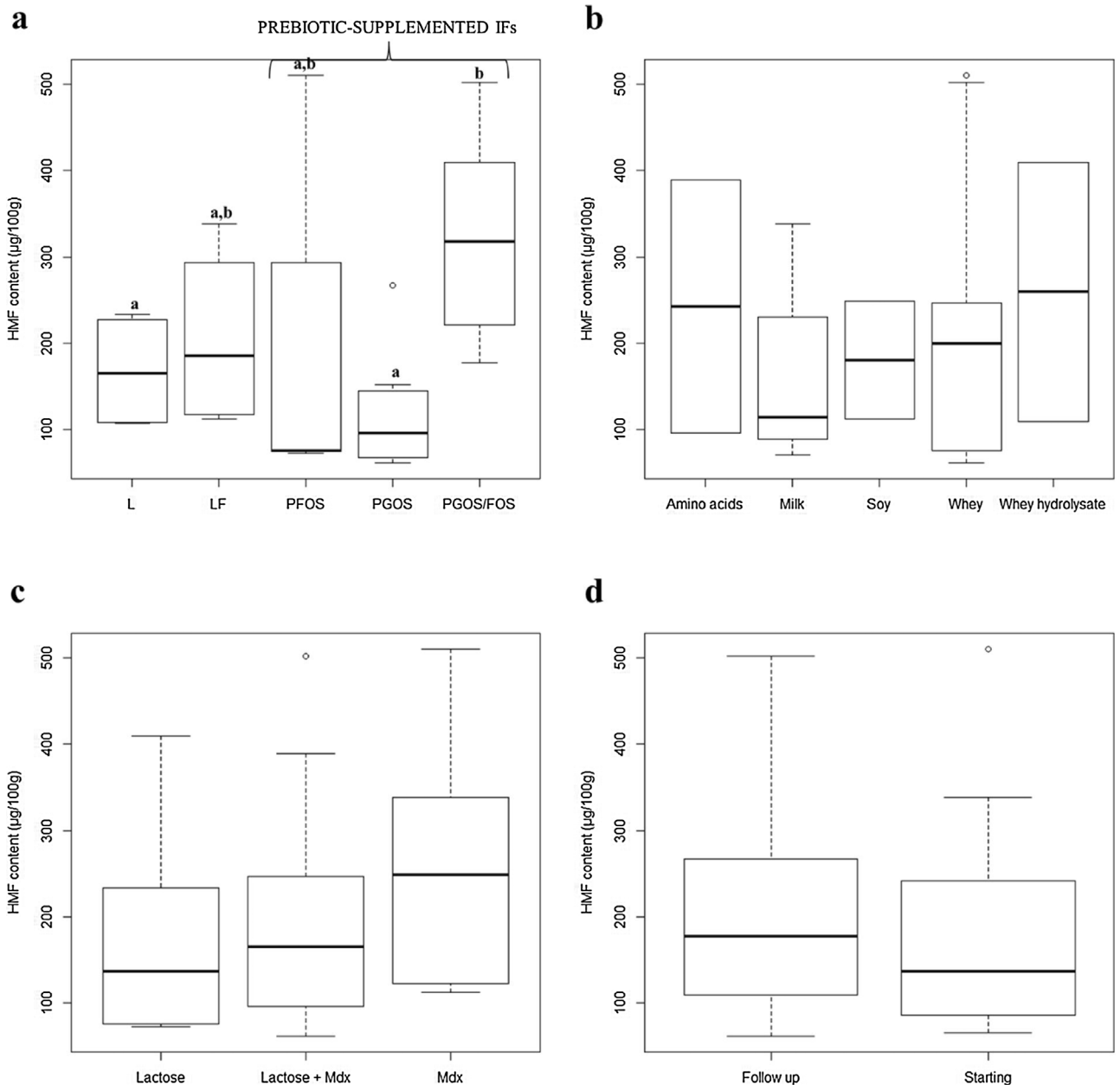


Fig. 2. Hydroxymethylfurfural (HMF) amounts found in infant formulas (IFs) **(a)** with lactose (L, $n = 4$), lactose-free (LF, $n = 4$)) and prebiotic-supplemented (with fructooligosaccharides, PFOS $n = 3$; with galactooligosaccharides, PGOS $n = 7$ and PGOS/FOS $n = 6$). **(b)** HMF levels found in IFs according to their protein source: amino acids ($n = 2$), milk ($n = 4$), soy ($n = 2$), whey ($n = 14$) and whey hydrolysate ($n = 2$). **(c)** HMF levels found in IFs according to their carbohydrate source: lactose ($n = 5$), maltodextrins (MdxMDX, $n = 5$) and lactose + MdxMDX ($n = 14$). **(d)** HMF levels found in IFs according infant age: follow-up ($n = 13$) and starting ($n = 11$). (Fig. 2a, ^{a,b} Statistically significant differences ($p < 0.05$) between IFs).

expressed as mean ± standard deviation (SD).

2.4. Determination of furosine

Furosine was determined by ion-pair RP-HPLC according to the method of Resmini et al. (1990). Before analysis, samples (0.250 mg) were hydrolysed with 4 mL of 8 N HCl at 110 °C for 24 h under inert conditions. The hydrolysate was filtered through Whatman No. 40 filter paper and 0.5 mL of filtrate were applied to a Sep-Pak C_{18} cartridge (Waters) previously activated with methanol and water. Furosine was eluted with 3 mL of 3 N HCl and 50 µL were injected into the chromatograph. RP-HPLC analysis was carried out using a C_8 column (250 mm × 4.6 mm, 5 µm particle size; Alltech furosine-dedicated, Nicholasville, KY) thermostated at 34 °C, using a linear binary gradient at a flow rate of 1.2 mL min^{-1}. Mobile phase consisted of solvent **A**, 0.4% acetic acid, and solvent **B**, 0.34% KCl in phase **A**. The elution program was as follows: 100% **A** from 0 to 8 min, 100% **B** from 8 to 21 min, and 100% **A** from 21 to 25 min. Detection was performed using a variable wavelength UV detector set at 280 nm (Agilent Technologies 1260 Infinity LC System − 1260 RID, Böblingen, Germany). Acquisition and processing of data were achieved with Agilent ChemStation Rev. C.01.05 software.

Calibration was performed by the external standard method using commercial furosine (Neosystem Laboratories, Strasbourg, France) as standard. All samples were analysed in duplicate, and data shown are the average values expressed as mg of furosine per 100 g of protein ± standard deviation (SD). The linearity of the method was evaluated with furosine standard in a concentration range of 0.023 to 6.990 mg L^{-1} and the obtained calibration curve was $y = 2.9378x + 0.0286$, with r^2 of 0.9995. Furthermore, limit of

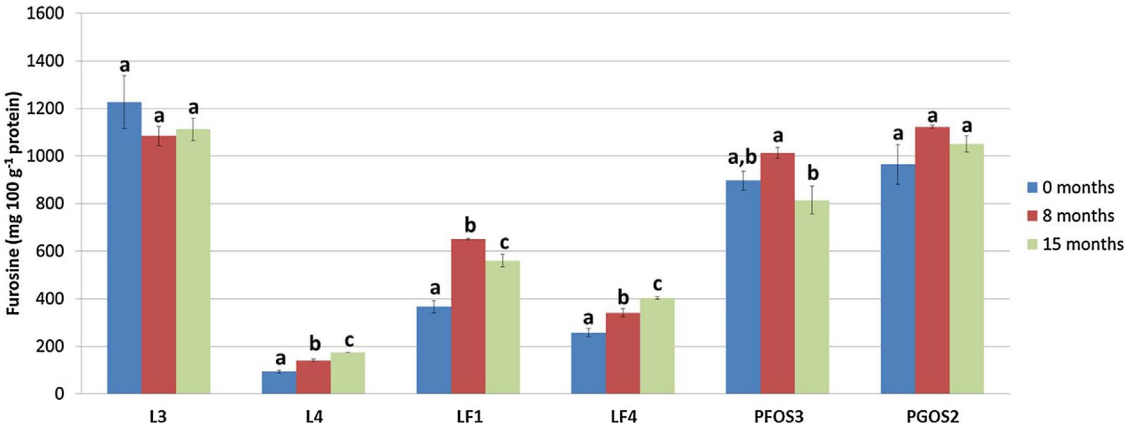


Fig. 3. Evolution of furosine content (mg 100 g^{-1} protein) during storage (8 and 15 months) at room temperature of six IFs: with lactose (L), lactose-free (LF) and prebiotic-supplemented IFs (PFOS and PGOS).[a,b,c] Statistically significant differences of furosine content during storage ($p < 0.05$).

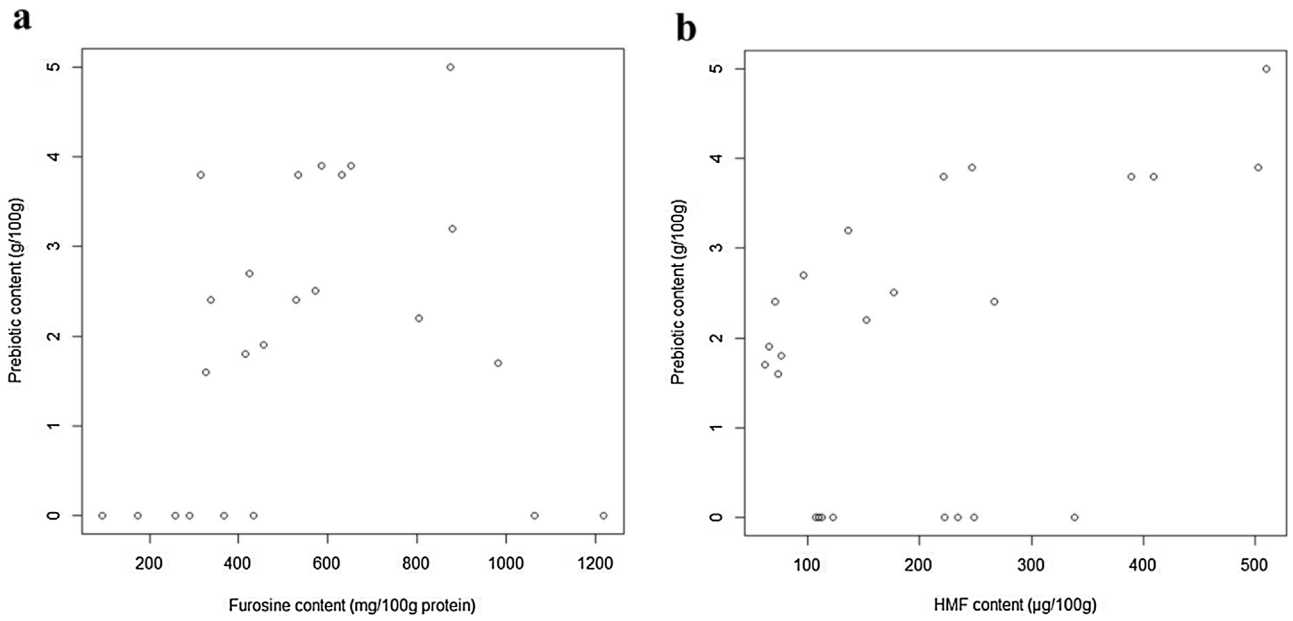


Fig. 4. Relationships between furosine (a) and HMF (b) and prebiotic content in studied IFs.

detection (0.018 mg L^{-1}) and quantification (0.060 mg L^{-1}) were established after the injection and analysis of the most diluted standard of furosine based on signal/noise ratios of 3:1 and 10:1, respectively (Teixido et al., 2006).

2.5. Determination of hydroxymethylfurfural (HMF)

The analysis of HMF was carried out by HPLC using an ACE-5C_{18} column (250 mm × 4.6 mm, 5 µm; Advanced Chromatography Technologies Ltd, Aberdeen, UK), thermostated at 25 °C and a linear gradient from methanol:water (5:95) to methanol:water (80:20) in 6 min, isocratic elution was then continued for 6 min and, finally, initial conditions were re-established in 1 min and held for 10 min. The flow rate was 1 mL min^{-1}mL/min and injection volume 50 µL. The UV detector was set at 283 nm (Rada-Mendoza et al., 2004).

Before chromatographic analysis, fat and protein interferences were removed from IFs by precipitation using Carrez reagents. IFs (400 mg) were gently mixed with 3 mL of water and equal volumes (0.4 mL) of Carrez I (7.2% w/v $K_4Fe(CN)_6{\cdot}3H_2O$ in water) and Carrez II (14% *w/v* $ZnSO_4{\cdot}7H_2O$ in water) in a 10-mL volumetric flask; 2 mL of methanol were added and the mixture was rised with water. Then, supernatant was collected, centrifuged at 10,000*g* for 5 min at room temperature, passed through a 0.45-µm filter (Waters) and then injected. Quantification was carried out employing the external standard method, using a commercial standard of HMF (Sigma, St. Louis, MO). Data were the mean values of duplicates expressed as mg 100 g^{-1} of product. The linearity of the method was evaluated with HMF standard in a concentration range of 0.4 to 400.0 µg L^{-1} (equivalent to 1 at 1000 µg 100 g^{-1} of product) and the obtained calibration curve was $y = 3.51x + 1.7836$, $r^2 = 0.9998$. Furthermore, limit of detection (0.3 µg L^{-1}, 0.75 µg 100 g^{-1} of product) and quantification (0.7 µg L^{-1}, 1.75 µg 100 g^{-1} of product) were determined.

2.6. Statistical analysis

ANOVA tests and Tukeý́s test at $p < 0.05$ were applied to data of furosine and HMF contents, in order to detect differences due to storage period (0, 8 and 15 months) and between the groups of IFs studied: IFs with and without prebiotics, starting and follow-up IFs, IFs containing proteins from different origin and different main carbohydrates. Box-and-whisker plots were used to summarise the data for furosine and HMF contents. In addition, Pearson correlation coefficient was calculated for all variables.

Multivariate data analysis was employed to investigate relationships among the 24 IFs with respect to quality parameters: general (pH, a_w, protein content), Maillard reaction (MR) indicators (furosine and HMF)

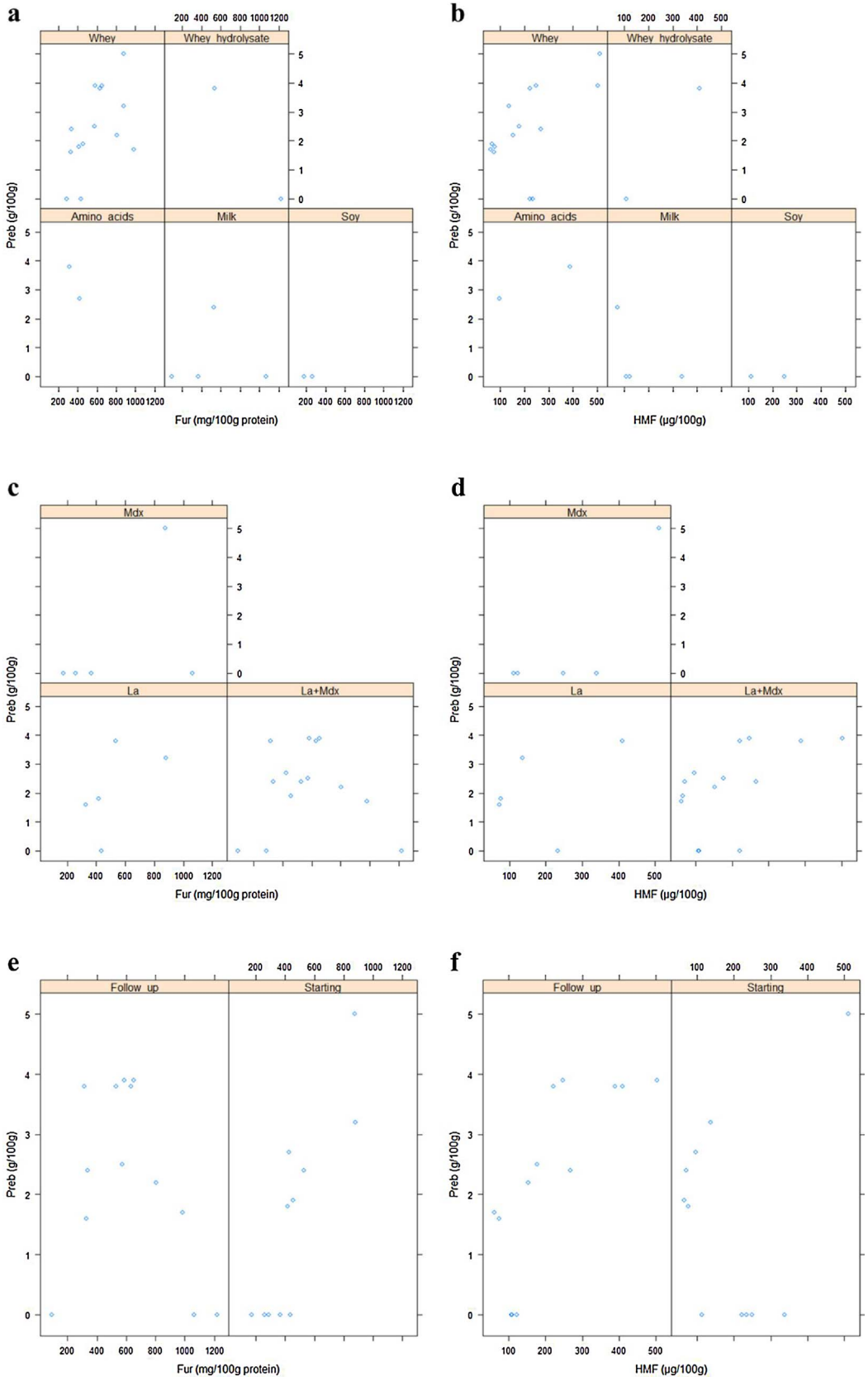


Fig. 5. Relationships between furosine (Fur) and HMF and prebiotic (Preb) content in studied IFs according to their protein (a, b) and carbohydrate (c, d) source and starting/follow-up type (e, f).

and carbohydrate contents, prebiotic and reducing sugars (considering the sum of monosaccharides, lactose, maltose and prebiotics) determined previously in our laboratory (Sabater et al., 2016). Two supervised classification methods were used, support vector machines (SVM) and random forests (RF). All variables were scaled and centered before the analysis. SVM classification algorithms try to find patterns in empirical data with regard to label classes (protein and carbohydrate source and IF type in this work). For SVM classification a radial basis function kernel was chosen.

In RF, a multitude of decision trees are constructed, outputting the different classes. Each node is split using the best among a subset of predictors randomly chosen. This method classifies very effectively compared to others, such as discriminant analysis, SVM and neural networks, and is robust against overfitting (Breiman, 2001).

These classification methods were validated by 10-fold cross-validation and tested with new samples. All statistical analyses were performed on R (R Core Team). SVM classification was performed with e1071 SVM package (Meyer et al., 2017). RF classification was performed with random Forest package (Liaw and Wiener, 2002).

3. Results and discussion

Data on physicochemical characteristics of analysed infant formulas (IFs) are shown in Table 1; as it can be observed, DM content and a_w were within the range of 94.5 to 99.2 and 0.07 to 0.28, respectively. In general, the composition of IFs was as expected for this type of product. Values of pH and protein content were in the range 6.67–7.23 and 8.31–16.0% (*w/w*), respectively. These values were fairly close to those found by Morales et al. (2004) and Chávez-Servín et al. (2015) in powdered infant formula.

3.1. Evaluation of Maillard reaction in commercial IFs

The content of furosine in commercial IFs is shown in Fig. 1. Furosine was detected in all studied IFs in variable amounts ranging from 94 to 1226 mg 100 g^{-1} protein, in samples without prebiotics and from 315 to 965 mg of furosine 100 g^{-1} of protein in prebiotic-supplemented IFs. No statistically significant differences between the presence and absence of prebiotics were found Fig. 1a. Morales et al. (2004) also obtained high values of furosine (271–1050 mg 100 g^{-1} protein) in Spanish commercial powdered IFs without prebiotics. Lactose-containing IFs (L) showed low levels, except L3, which presented 1226 mg of furosine 100 g^{-1} of protein; this fact could be attributed to the presence of whey protein hydrolysate, which contains high levels of free amino acids that may participate in the Maillard reaction (MR). Contreras-Calderón et al. (2008) found high values of furosine (644–1435 mg 100 g^{-1} of protein) in whey proteins with different lactose contents used as ingredients in IFs, which possessed a high available lysine content. Taking into account prebiotic-enriched IFs, those containing galactooligosaccharides (PGOS) showed the highest amount of furosine i.e. 965 mg 100 g^{-1} of protein, while among fructooligosaccharide enriched formulas, PFOS3 had the highest content (897 mg 100 g^{-1} of protein).

As is well known, the final quality of IFs depends on formulation and processing design, both being influential factors in the evolution of MR (Contreras-Calderón et al., 2009). Therefore, sample PGOS5, with similar formulation to that of PGOS2, presented a higher furosine level, which could be attributed to higher temperature conditions during processing or storage.

Finally, the significant differences in furosine content within the three groups of IFs (i.e considering source of protein, main carbohydrates and starting/follow-up types) were studied using two-way ANOVA. Significant differences depending on source of protein were observed, especially between IFs with soy protein and IFs with hydrolysed whey protein, confirmed by a pairwise test (Fig. 1b); however, no other differences were found (Fig. 1c; d).

In general, the furosine values determined in IFs were in the range found by Contreras-Calderón et al. (2009), Contreras-Calderón et al. (2017), Fenaille et al. (2006), Ferrer et al. (2003), and Martysiak-Zurowska and Stolyhwo (2007) in conventional, hypoallergenic and soybean-based IFs. However, other studies reported lower furosine contents in this type of product (Michalak et al., 2006). Therefore it is necessary to control the heat treatment applied to IFs because high furosine values can indicate advanced degradation of proteins, resulting in a considerable decrease of lysine in IFs and therefore a decrease in nutritive value (Contreras-Calderón et al., 2017).

The content of HMF in all analysed IFs is shown in Fig. 2a. Free HMF values ranged from 62 to 510 μg 100 g^{-1} of product, and only statistically significant differences between PGOS/FOS IFs and conventional (L) and PGOS IFs were found. On the other hand, there were no significant differences between IFs containing: proteins from different origins (Fig. 2b), different main carbohydrate (Fig. 2c), nor starting and follow-up IFs (Fig. 2d). These results agree with those obtained by Chávez-Servín et al. (2006) in milk-based IFs and Contreras-Calderón et al. (2017) in adapted, follow-up, hypoallergenic and soybean-based IFs. Nevertheless, free HMF contents were lower than those reported in several studies for powdered IFs (Albalá-Hurtado et al., 1997; Guerra-Hernández et al., 2002a; Michalak et al., 2006; Kocadağlı and Gökmen, 2014). Differences in the HMF levels could be attributable to IFs composition, thermal treatments during manufacturing and storage.

3.2. Progress of Maillard reaction in IFs during storage

The influence of IFs storage (8 and 15 months), at room temperature, on MR progress has also been studied in some representative samples; the results are depicted in Fig. 3. Furosine content increased significantly ($p < 0.05$) in lactose-containing IF L4 and lactose-free IF LF4 (with soy protein) at the end of the storage period; these samples have a low initial level of furosine. In contrast, it can be observed that furosine levels found in two samples (LF1, PFOS3) were lower in the 15th month of storage. These results suggest that furosine may suffer degradation, since other compounds from more advanced MR stages are formed (Ferrer et al., 2003). Ferrer et al. (2003), Gliguem and Birlouez-Aragon (2005) and Guerra-Hernández et al. (2002a) studied the evolution of MR products throughout the shelf-life storage period and observed an increase of furosine content which was statistically significant. On the other hand, no changes during storage were found for HMF, with the exception of FOS-enriched formula PFOS3, where HMF content increased, in accordance with the furosine decrease after 15 months. Chávez-Servín et al. (2006) found no increase in free HMF in milk-based IFs after 12 months of storage. However, these results differ from those reported by Chávez-Servín et al. (2005), who studied the evolution over shelf life of HMF levels in milk-based formulas from production until 15 months. They found a significant increase of free HMF after 15 months, and in another study (Chávez-Servín et al., 2015) determined slight increases of furfural contents in powdered IFs after 70 days of storage at room temperature. Ferrer et al. (2005) also reported an increase in the HMF content of milk-based IFs at the end of the storage during 24 months at 20 °C.

3.3. Multivariate data analysis

In order to obtain more information on MR development in IFs here studied, the relationships between furosine and HMF and the different parameters determined in this paper (pH, a_w and protein content), as well as those published by Sabater et al. (2016) (reducing sugars and prebiotic contents), calculating the Pearson correlation coefficient, have been studied. The results have shown a strong correlation between HMF and prebiotic content (correlation coefficient 0.49) (Fig. 4b), HMF increasing with prebiotic content. However, a great dispersion of furosine values were observed (Fig. 4a) and no significant correlation with prebiotic content (correlation coefficient 0.21) was obtained.

The second step was to apply two classification methods to study the relationship between three previously established categories of IFs according to i) protein source, ii) carbohydrate source and iii) starting/follow-up type, and the set of quality parameters studied, general (pH, a_w, protein), related with carbohydrate composition (reducing sugars and prebiotic content) (Sabater et al., 2016), and with MR indicators, furosine and HMF. The supervised models, trained with 70% IFs (including samples stored for 8 months) and tested with 30% IFs, were applied. As can be seen using an SVM classification method, a high classification rate was obtained (categories shown in Fig. 5): a) and b) protein source training, validation and test rates, 95.0, 92.5 and 90.0% (cost value 2.0, gamma value 0.5); c) and d) carbohydrate source training, validation and test rates, 100, 90.0 and 90.0% (cost value 1.0, gamma value 0.5); and e) and f) starting/follow-up IF training, validation and test rates, 100, 92.5 and 90.0% (cost value 2.0, gamma value 0.5). Cost parameter indicates how much the misclassification of samples should be avoided during the SVM application. A large cost gives high variance and low bias because the cost of misclassification is highly penalised. A small cost gives higher bias and lower variance. Gamma parameter defines how far the influence of a single training example reaches, with low values meaning a high influence. Therefore, a small gamma value gives low bias and high variance while a large gamma value gives higher bias and low variance. Therefore, it can be said that good classifications are obtained. With regard to variable importance, furosine and reducing sugars content showed the highest coefficients (0.41 and 0.27, respectively) when IFs were classified according to their protein source. When IFs were classified according to their carbohydrate source, reducing sugars showed the highest coefficient, 0.60. Finally, when IFs were classified according to their type, furosine and protein content showed the highest coefficients (0.34 and 0.29, respectively). In all three cases, the models were able to accurately predict new samples.

On the other hand, to corroborate the good fit classification reached when SVM was used, RF was also applied to classify IFs, and similar results were obtained. According to their protein source, the training, validation and test rates were 90, 89 and 80%. This RF classification model was built with 150 trees and 1 variable tried at each split. With regard to variable importance in this classification, prebiotic and reducing sugars content showed the highest Mean Decrease Accuracy (10.0 and 9.1, respectively) while reducing sugars and furosine content showed the highest Mean Decrease Gini (3.8 and 3.6, respectively). These two parameters represent the loss of accuracy and the mean decrease in node purity when a variable is removed. pH, protein and reducing sugars content were the most influent variables to classify IFs containing amino acids (coefficients −1.0, −1.7, 1.0, respectively). pH and reducing sugars content were the most influent variables to classify milk IFs (coefficients 8.2 and 6.3, respectively). Furosine and prebiotic content were the most influent variables to classify soy IFs (coefficients 7.0 and 5.3, respectively). HMF and prebiotic content were the most influent variables to classify IFs containing whey (coefficients 7.5 and 7.5, respectively). Protein content was the most influent variable to classify IFs containing whey hydrolysate (coefficient 5.3).

When RF were applied to classify IFs according to their carbohydrate source, the training, validation and test rates were 95, 92 and 90%. This RF classification model was built with 100 trees and 1 variable tried at each split. With regard to variable importance in this classification, reducing sugars content showed the highest Mean Decrease Accuracy and Mean Decrease Gini (8.9 and 6.5, respectively). Protein and a_w were the most influent variables to classify lactose IFs (coefficients 5.5 and 5.2, respectively). Reducing sugars and HMF content were the most influent variables to classify IFs containing lactose and maltodextrins (coefficients 6.1 and 4.0, respectively). Reducing sugars and HMF content were the most influent variables to classify IFs containing maltodextrins (coefficients 9.3 and 5.1, respectively).

Finally, when RF was applied to classify IFs according to starting/follow-up, the training, validation and test rates were 97.5, 92 and 85%. This RF classification model was built with 100 trees and 3 variables tried at each split. With regard to variable importance in this classification, furosine and reducing sugars content showed the highest Mean Decrease Accuracy (9.4 and 8.1, respectively) and highest Mean Decrease Gini (5.7 and 5.3, respectively). Reducing sugars and furosine content were the most influent variables to classify IFs by infant ages, the coefficients being 8.2 and 7.8 for starting and 6.6 and 7.3 for follow-up IFs.

RF reinforce results from SVM. In general, SVM classification performance was slightly better than RF classification. Although, in both cases, these results indicate that it is possible to classify the IFs according to the protein or carbohydrate source and starting/follow-up type, considering the set of quality parameters determined in this study, which could be applied to IFs of unknown origin for classification, considering the high prediction rates of new samples.

4. Conclusion

The results here obtained show thermal damage of proteins of different types of IFs through measurement of furosine (2-FM-Lys) and HMF. Although great variability in furosine and HMF contents was found in the analysed commercial IFs, no significant differences were found in furosine content, taking into account factors such as presence of prebiotics, main carbohydrates present or starting or follow-up IFs. With respect to HMF content, scarce differences between IFs with and without prebiotics were observed. In addition, storage of IFs at room temperature did not produce important changes in the furosine and HMF contents. Therefore, the high levels of furosine detected in some IFs may be attributed to excessive heat treatment during processing, elevated temperature during storage or presence of high proportion of furosine in some of the ingredients used in IFs elaboration. The use of furosine and HMF as indicators of thermal damage, determines the quality of prebiotic-supplemented IFs, which is similar to IFs without added prebiotics. Finally, the utility of the set of quality parameters here determined to classify IFs using appropriate statistical tools (SVM and RF) has been established. IFs were classified according to the protein or carbohydrate source and starting/follow-up type using these supervised models, which could be applied to samples of unknown origin.

Acknowledgements

This work has been supported by the Spanish MINECO (AGL2014-53445-R), ALIBIRD-CM S-2013/ABI-272 and Spanish Danone Institute. Carlos Sabater thanks his FPU Predoctoral contract from Spanish MECD (FPU14/03619).

References

Association of Official Analytical Chemists (AOAC), 1990a. Official Methods of Analysis of the AOAC. Method 920.151, 15th ed. EEUU: Association of Official Analytical Chemists, Arlington, VA.

Association of Official Analytical Chemists (AOAC), 1990b. Official Methods of Analysis of the AOAC. Method 962.18, 15th ed. Association of Official Analytical Chemists, Arlington, VA.

Ackerman, D.L., Craft, K.M., Townsend, S.D., 2017. Infant food applications of complex carbohydrates: structure synthesis, and function. Carbohydr. Res. 437, 16–27.

Albalá-Hurtado, S., Veciana-Nogués, M.T., Izquierdo-Pulido, M., Vidal-Carou, M.C., 1997. Determination of free and total furfural compounds in infant milk formulas by high-performance liquid chromatography. J. Agric. Food Chem. 45 (6), 2128–2133.

Ben, X.M., Li, J., Feng, Z.T., Shi, S.Y., Lu, Y.D., Chen, R., et al., 2008. Low level of galacto-oligosaccharide in infant formula stimulates growth of intestinal Bifidobacteria and Lactobacilli. World J. Gastroenterol. 14 (42), 6564–6568.

Boitz, L.I., Mayer, H.K., 2016. Analytical assessment of the intense heat load of whipping cream, coffee cream, and condensed milk at retail in Austria and Germany. Dairy Sci. Technol. 96, 677–692.

Breiman, L., 2001. Random forests. Mach.Learn. 45 (1), 5–32.

Chávez-Servín, J.L., Castellote, A.I., López-Sabater, M.C., 2005. Analysis of potential and free furfural compounds in milk-based formulae by high-performance liquid

chromatography: evolution during storage. J. Chromatogr. A 1076 (1), 133–140.
Chávez-Servín, J.L., Castellote, A.I., López-Sabater, M.C., 2006. Evolution of potential and free furfural compounds in milk-based infant formula during storage. Food Res. Int. 39 (5), 536–543.
Chávez-Servín, J.L., de la Torre Carbot, K., García-Gasca, T., Castellote, A.I., López-Sabater, M.C., 2015. Content and evolution of potential furfural compounds in commercial milk-based infant formula powder after opening the packet. Food Chem. 166, 486–491.
Contreras-Calderón, J., Guerra-Hernández, E., García-Villanova, B., 2008. Indicators of non-enzymatic browning in the evaluation of heat damage of ingredient proteins used in manufactured infant formulas. Eur. Food Res. Technol. 227 (1), 117–124.
Contreras-Calderón, J., Guerra-Hernández, E., García-Villanova, B., 2009. Utility of some indicators related to the Maillard browning reaction during processing of infant formulas. Food Chem. 114 (4), 1265–1270.
Contreras-Calderón, J., Guerra-Hernández, E., García-Villanova, B., Gómez-Narváez, F., Zapata-Betancur, A., 2017. Effect of ingredients on non-enzymatic browning: nutritional value and furanic compounds in spanish infant formulas. J. Food Nutr. Res. 5 (4), 243–252.
Corzo-Martínez, M., Sánchez, C.C., Moreno, F.J., Patino, J.M.R., Villamiel, M., 2012. Interfacial and foaming properties of bovine β-lactoglobulin: galactose Maillard conjugates. Food Hydrocolloids 27 (2), 438–447.
Fabris, A., Biasioli, F., Granitto, P.M., Aprea, E., Cappellin, L., Schuhfried, E., Soukoulis, C., Märk, T.D., Gasperi, F., Endrizzi, I., 2010. PTR-TOF-MS and data-mining methods for rapid characterisation of agro-industrial samples: influence of milk storage conditions on the volatile compounds profile of Trentingrana cheese. J. Mass Spectrom. 45 (9), 1065–1074.
Fanaro, S., Boehm, G., Garssen, J., Knol, J., Mosca, F., Stahl, B., et al., 2005. Galactooligosaccharides and long-chain fructo-oligosaccharides as prebiotics in infant formulas: a review. Acta Paediatr. 94 (s449), 22–26.
Fenaille, F., Parisod, V., Visani, P., Populaire, S., Tabet, J.C., Guy, P.A., 2006. Modifications of milk constituents during processing: a preliminary benchmarking study. Int. Dairy J. 16 (7), 728–739.
Ferrer, E., Alegría, A., Farré, R., Abellán, P., Romero, F., Clemente, G., 2003. Evolution of available lysine and furosine contents in milk-based infant formulas throughout the shelf-life storage period. J. Sci. Food Agric. 83 (5), 465–472.
Ferrer, E., Alegria, A., Farré, R., Abellán, P., Romero, F., 2005. High-performance liquid chromatographic determination of furfural compounds in infant formulas during full shelf-life. Food Chem. 89 (4), 639–645.
Fratianni, A., Niro, S., Messia, M.C., Cinquanta, L., Panfili, G., Albanese, D., Di Matteo, M., 2017. Kinetics of carotenoids degradation and furosine formation in dried apricots (Prunus armeniaca L.). Food Res. Int. 99, 862–867.
Gliguem, H., Birlouez-Aragon, I., 2005. Effects of sterilization, packaging, and storage on vitamin C degradation, protein denaturation, and glycation in fortified milks. J. Dairy Sci. 88 (3), 891–899.
Gokmen, V., Serpen, A., Acar, O.C., Morales, F.J., 2008. Significance of furosine as heat-induced marker in cookies. J. Cereal Sci. 48 (3), 843–847.
Guerra-Hernández, E., Leon, C., Corzo, N., García-Villanova, B., Romera, J.M., 2002a. Chemical changes in powdered infant formulas during storage. Int. J. Dairy Technol. 55 (4), 171–176.
Guerra-Hernández, E., Leon Gomez, C., Garcia-Villanova, B., Corzo Sanchez, N., Romera Gomez, J.M., 2002b. Effect of storage on non-enzymatic browning of liquid infant milk formulae. J. Sci. Food Agric. 82 (5), 587–592.
Kim, J.S., 2010. Influence of the pH and enantiomer on the antioxidant activity of Maillard Reaction mixture solution in the model systems. J. Food Nutr. 15, 287–296.
Knol, J., Scholtens, P., Kafka, C., Steenbakkers, J., Gross, S., Helm, K., et al., 2005. Colon microflora in infants fed formula with galacto- and fructo-oligosaccharides: more like breast-fed infants. J. Pediatr. Gastroenterol. Nutr. 40 (1), 36–42.
Kocadağlı, T., Gökmen, V., 2014. Investigation of α-dicarbonyl compounds in baby foods by high-performance liquid chromatography coupled with electrospray ionization mass spectrometry. J. Agric. Food Chem. 62 (31), 7714–7720.
Liaw, A., Wiener, M., 2002. Classification and regression by random forest. R News 2 (3), 18–22.
Martysiak-Zurowska, D., Stolyhwo, A., 2007. Content of furosine in infant formulae and follow-on formulae. Pol. J. Food Nutr. Sci. 57 (2), 185–190.
Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., 2017. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071). TU Wien R package version 1. 6–8. https://CRAN.R-project.org/package=e1071.
Michalak, J., Kuncewicz, A., Gujska, E., 2006. Monitoring selected quality indicators of powdered infant milk formulas. Pol. J. Food Nutr. Sci. 15 (56), 131–135.
Morales, F.J., Romero, C., Jimenez-Perez, S., 1996. Evaluation of heat-induced changes in Spanish commercial milk: hydroxymethylfurfural and available lysine content. Int. J. Food Sci. Technol. 31 (5), 411–418.
Morales, V., Olano, A., Corzo, N., 2004. Ratio of maltose to maltulose and furosine as quality parameters for infant formula. J. Agric. Food Chem. 52 (22), 6732–6736.
Nasirpour, A., Scher, J., Desobry, S., 2006. Baby foods: formulations and interactions (a review). Crit. Rev. Food Sci. Nutr. 46 (8), 665–681.
R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria URL. https://www.R-project.org/.
Ríos-Ríos, K.L., Vázquez-Barrios, M.E., Gaytán-Martínez, M., Olano, A., Montilla, A., Villamiel, M., 2018. 2-Furoylmethyl amino acids as indicators of Maillard reaction during the elaboration of black garlic. Food Chem. 240, 1106–1112.
Rada-Mendoza, M., Sanz, M.L., Olano, A., Villamiel, M., 2004. Formation of hydroxymethylfurfural and furosine during the storage of jams and fruit-based infant foods. Food Chem. 85 (4), 605–609.
Resmini, P., Pellegrino, L., Battelli, G., 1990. Accurate quantification of furosine in milk and dairy products by a direct HPLC method. Ital. J. Food Sci. 2 (3), 173–183.
Roux, S., Courel, M., Birlouez-Aragon, I., Municino, F., Massa, M., Pain, J.-P., 2016. Comparative thermal impact of two UHT technologies, continuous ohmic heating and direct steam injection, on the nutritional properties of liquid infant formula. J. Food Eng. 179, 36–43.
Sabater, C., Prodanov, M., Olano, A., Corzo, N., Montilla, A., 2016. Quantification of prebiotics in commercial infant formulas. Food Chem. 194, 6–11.
Teixido, E., Santos, F.J., Puignou, L., Galceran, M.T., 2006. Analysis of 5-hydroxymethylfurfural in foods by gas chromatography-mass spectrometry. J. Chromatogr. A 1135 (1), 85–90.
Wu, D., Feng, S., He, Y., 2008. Short-wave near-infrared spectroscopy of milk powder for brand identification and component analysis. J. Dairy Sci. 91 (3), 939–949.

Food Microbiology 77 (2019) 93–105

Contents lists available at ScienceDirect

Food Microbiology

journal homepage: www.elsevier.com/locate/fm

Synthesis of prebiotic galactooligosaccharides from lactose and lactulose by dairy propionibacteria

Carlos Sabater[a], Agustina Fara[b], Jorge Palacios[b], Nieves Corzo[a], Teresa Requena[c], Antonia Montilla[a], Gabriela Zárate[b,*]

[a] Grupo de Química y Funcionalidad de Carbohidratos y Derivados, CIAL (CSIC-UAM), Calle Nicolás Cabrera, 9, 28049, Madrid, Spain
[b] Laboratorio de Ecofisiología Tecnológica, CERELA-CONICET, Chacabuco 145, T4000ILC, San Miguel de Tucumán, Argentina
[c] Grupo de Biología Funcional de Bacterias Lácticas, CIAL (CSIC-UAM), Calle Nicolás Cabrera, 9, 28049, Madrid, Spain

ARTICLE INFO

Keywords:
Propionibacteria
β-galactosidase
Transgalactosylation
Oligosaccharides
Prebiotics

ABSTRACT

The potential of probiotic bacteria to produce prebiotic oligosaccharides by transgalactosylation has been minimally studied. In this work, we screened the β-galactosidase (β-gal) activity of dairy propionibacteria (PAB) isolated from Argentinean foods to select strains for the synthesis of oligosaccharides from lactose (GOS) and lactulose (OsLu). PAB, when grown in a medium with lactose as a carbon source, were disrupted, and the cell-free extracts were assayed for β-gal activity. Nine strains grew on lactose and showed β-gal activities from 0.27 to 2.60 U mL^{-1}. *Propionibacterium acidipropionici* LET 120, the strain with the highest activity, was able to synthesize, using 30% lactose and lactulose at pH 6.5 and 45 °C, 26.8% of LET 120-GOS and 26.1% of LET 120-OsLu after 24 h. When they were tested as carbon sources for growth, *P. acidipropionici* LET 120 attained higher biomasses, μ_{max} and β-gal activities at the expense of *Aspergillus oryzae*-OsLu, Vivinal®-GOS and lactulose compared to lactose or glucose. In addition, LET 120-GOS and LET 120-OsLu synthesized by PAB were prebiotic for some probiotic strains. For the first time, our results show the production of GOS and OsLu by dairy PAB, and these results encourage further studies on the optimization of the synthesis and structure characterization of the obtained oligosaccharides.

1. Introduction

In the last several decades, functional foods, i.e., those that supply health benefits beyond basic nutrition, have received great attention from the food industry due to the high demand of healthy products by consumers worldwide. In this respect, probiotic microorganisms and prebiotic compounds are the functional ingredients most widely studied and the only two recognized by Argentinean legislation and the National Food Code of the country (Código Alimentario Argentino, 1969, www.anmat.gov.ar).

Probiotics are live microorganisms that when administered in adequate amounts improve host health through different mechanisms (Hill et al., 2014). One of the first beneficial effects ascribed to probiotics is their ability to contribute to lactose metabolism in the gut and the alleviation of intolerance symptoms. This effect is mediated by the β-galactosidase (β-gal) enzyme (EC 3.2.1.23) of the microorganisms that are included in the probiotic product and which hydrolyzes lactose in food before consumption and *in situ* in the intestine (Vonk et al., 2012). However, β-gal have more biotechnological applications, and depending on the reaction conditions (substrate concentration, a_w, pH, temperature, time), the enzyme may display hydrolase or transferase activities. In particular and due to transferase activity, there is a production of prebiotic oligosaccharides by the transgalactosylation of galactose moieties to a carbohydrate acceptor instead of water (Moreno et al., 2014; Vera et al., 2016; Zárate et al., 2013). Prebiotics are defined as "non-digestible food ingredients that beneficially affect host health by selectively stimulating the growth and/or activity of one or a limited number of desirable bacteria in the body" (Gibson et al., 2017). Among the prebiotics, fructans (inulin and fructooligosaccharides), galactooligosaccharides (GOS) and lactulose are recognized as useful for the modulation of colonic microbiota toward a healthy balance, which usually involves the increase of bifidobacteria and lactobacilli at the expense of less desirable organisms such as Clostridia, Bacteroides and Enterobacteria (Cardelle-Cobas et al., 2009a; Roberfroid et al., 2010).

As mentioned, some prebiotics, such as GOS, lactulose and its derived oligosaccharides (OsLu), can be obtained enzymatically by using microbial β-gal (Cardelle Cobas et al., 2008a, 2008b; Guerrero et al.,

* Corresponding author.
E-mail address: gzarate@cerela.org.ar (G. Zárate).

https://doi.org/10.1016/j.fm.2018.08.014
Received 9 March 2018; Received in revised form 28 July 2018; Accepted 27 August 2018
Available online 27 August 2018
0740-0020/

2011; Martínez Villaluenga et al., 2008a, 2008b). However, depending on the origin of β-gal and reaction conditions, the amount of GOS obtained, composition of monomers and linkages between D-galactose units may vary, affecting their properties and prebiotic potential. Previous studies have shown the synthesis, structural characterization and prebiotic effect of GOS and OsLu produced by fungal β-gal from *Aspergillus oryzae* (Cardelle-Cobas et al., 2016; Urrutia et al., 2013; Vera et al., 2012), *A. aculeatus* and *Kluyveromyces lactis* (Cardelle-Cobas et al., 2009b, 2008a; 2008b; Hernández-Hernández et al., 2011; Martínez-Villaluenga et al., 2008b). In the same manner, GOS synthesized by bacteria such as *Bacillus circulans* (Corzo Martínez et al., 2013; Yin et al., 2017), *Lactobacillus reuteri* (Splechtna et al., 2006), *L. plantarum* (Benavente et at., 2015; Iqbal et al., 2010) and *Bifidobacterium* species (Hsu et al., 2007; Osman et al., 2010) have been reported.

Dairy propionibacteria (PAB) are microorganisms that are traditionally used by industry for the manufacture of Swiss-type cheeses and the biologic production of propionic acid. However, research on their potential application to the production of relevant biomolecules, such as vitamins, conjugated linoleic acid (CLA), exopolysaccharides (EPS), trehalose, and bacteriocins, has significantly increased in recent years. In addition, several probiotic effects have been reported for dairy PAB, which could be due to their ability to modulate gut physiology, microbiota composition and host immunity in a beneficial manner (Rabah et al., 2017; Zárate and Pérez Chaia, 2015). As result, *Propionibacterium freudenreichii* and *Propionibacterium acidipropionici* have been included in the list of agents recommended for the Qualified Presumption of Safety (QPS) of the European Food Safety Authority (EFSA, 2013).

Regarding lactose metabolism, both *P. freudenreichii* and *P. acidipropionici* hydrolyze lactose by β-gal activity since no phospho-β-gal was detected (Hartley and Vedamuthu, 1975; Zárate et al., 2003). The main biochemical characteristics of the enzyme and its regulation in the presence of lactose and lactate were determined to decide on an appropriate vehicle to deliver PAB to the host (Zárate and Pérez Chaia, 2012; Zárate et al., 2003). Since β-gal of *P. acidipropionici* was able to resist both the manufacture of a Swiss-type cheese and gastrointestinal conditions, it could contribute to lactose digestion included in a probiotic product (Pérez Chaia and Zárate, 2005; Zárate et al., 2000). Mice fed with *P. acidipropionici* CRL 1198 that was included in a milk or cheese, showed increased levels of β-gal in the small bowel and high propionic acid concentrations in the caecum that could favor the recovery of water and electrolytes involved in osmotic diarrhea induced by unabsorbed lactose (Pérez Chaia and Zárate, 2005). Although β-gal from dairy PAB has shown properties for probiotic purposes, its transglycosidase activity for prebiotic synthesis was not investigated. In the present study, we assessed the ability of PAB isolated from Argentinian dairy products to produce GOS and OsLu and their potential prebiotic activity.

2. Materials and methods

2.1. Chemical and reagents

All chemicals and reagents used were of analytical grade (glucose, galactose, fructose, lactose, lactulose, raffinose, and stachyose used as standards and enzyme substrates) and were purchased from Sigma-Aldrich (St. Louis, MO, USA). Acetonitrile of high-performance liquid chromatography (HPLC) grade was purchased from Merck (Darmstadt, Germany). The commercial preparation Lactozym®Pure 6500L (Lactozym) was a gift from Novozymes (Dittingen, Switzerland) and was used as enzyme control. Vivinal®-GOS syrup (Friesland Campina Domo, The Netherlands) was kindly provided by Dr. A. Illanes (EIB, PUCV, Chile) and it contained 75% dry matter (DM), and a carbohydrate composition of 59% of GOS, 19% of lactose, 21% of glucose and 1.4% of galactose. OsLu were synthesized using a commercial lactulose preparation (670 g L^{-1}; Duphalac®, Abbott Biologicals B.V., Olst, The Netherlands) and β-gal from *Aspergillus oryzae* (16 U mL^{-1}; Sigma, St. Louis, MO) according to the protocol described in Algieri et al. (2014). The analysis of OsLu showed a DM of 81% and a carbohydrate composition of 14% of galactose, 4% lactose, 24% lactulose, and 58% OsLu (23% disaccharides, 28% trisaccharides and 8% tetrasaccharides (López-Sanz et al., 2015). MRS culture medium for microbial growth was purchased from Pronadisa (Madrid, Spain) and Britania (Buenos Aires, Argentina).

2.2. Microorganisms and culture conditions

Seventeen strains belonging to *P. freudenreichii, P. jensenii* and *P. acidipropionici* isolated in CERELA-CONICET from raw milk and Swiss-type cheeses made in Argentina were used in this study. *P. freudenreichii* subsp. *shermanii* ATCC 13673 and *P. acidipropionici* ATCC 25562 from American Type Culture Collection were used as a reference (Table 1). *P. acidipropionici* LET 120 was molecularly identified by sequencing the 16S rDNA as described by Lorenzo Pisarello et al. (2010). The obtained sequence was submitted to the European Nucleotide Archive database (Accession number PRJEB25502). *Bifidobacterium animalis* subsp. *lactis* BB-12®, *Lactobacillus casei* CRL 431® (from Chr.Hansen) and *Escherichia coli* C3 (from the Institute of Microbiology "Luis Verna" of the University of Tucumán) were used to test the prebiotic activity of oligosaccharide mixtures. The strains were stored at −20 °C in 10% (w/v) reconstituted skim milk (RSM) containing 5 g L^{-1} of yeast extract and 10% glycerol and were activated by three successive transfers every 24 h in MRS broth or LB broth for *E. coli*.

2.3. Assessment of the β-galactosidase activity of dairy propionibacteria (PAB)

The β-gal activity of PAB was determined in cell-free extracts (CFE)

Table 1
Growth parameters and β-gal production by dairy PAB developed after 24 h of incubation at 37 °C on MRS containing lactose as carbon source. Data are the means of two independent assays (± SD). The r^2 for fit of data to sigmoid curves ranged from 0.985 to 0.998 for all the strains.

Strain	Maximum CFU mL^{-1}	Growth rate (μ; h^{-1})	Final pH	Activity (U mL^{-1})	Specific activity (U mg^{-1})
*P**. *acidipropionici* LET 113	2.50 x $10^{9a,b}$ (0.11)	$0.09^{a,b}$ (0.01)	$4.80^{a,b}$ (0.04)	$0.680^{a,b}$ (0.131)	$0.283^{a,b}$ (0.054)
P. acidipropionici LET 116	2.26 x $10^{9a,b,c}$ (0.05)	$0.08^{a,b,c}$ (0.00)	$4.85^{a,b}$ (0.04)	0.270^{b} (0.055)	0.085^{b} (0.017)
P. acidipropionici LET 117	2.16 x $10^{9a,b,c,d}$ (0.01)	$0.09^{a,b}$ (0.00)	$4.81^{a,b}$ (0.06)	$0.380^{a,b}$ (0.067)	0.081^{b} (0.014)
P. acidipropionici LET 119	2.29 x $10^{9a,b,c}$ (0.04)	$0.09^{a,b,c}$ (0.01)	$4.80^{a,b}$ (0.08)	$0.720^{a,b}$ (0.195)	0.104^{b} (0.028)
P. acidipropionici LET 120	3.01 x 10^{9a} (0.08)	0.10^{a} (0.01)	4.75^{b} (0.04)	2.593^{c} (0.156)	2.553^{c} (0.153)
P. acidipropionici ATCC 25562**	2.06 × $10^{9b,c,d}$ (0.20)	$0.08^{a,b,c}$ (0.01)	$4.80^{a,b}$ (0.03)	1.950^{d} (0.077)	0.641^{e} (0.025)
P. freudenreichii LET 114	1.41 x $10^{9c,d,e}$ (0.04)	$0.07^{b,c}$ (0.01)	$4.95^{a,b}$ (0.10)	$0.670^{a,b}$ (0.220)	$0.214^{a,b}$ (0.070)
P. freudenreichii LET 125	1.31 × $10^{9d,e}$ (0.22)	$0.06^{c,d}$ (0.01)	$4.90^{a,b}$ (0.11)	$0.676^{a,b}$ (0.133)	$0.248^{a,b}$ (0.049)
P. freudenreichii LET126	1.08 x 10^{9e} (0.06)	$0.07^{b,c}$ (0.01)	$5.01^{a,b}$ (0.11)	$0.627^{a,b}$ (0.064)	$0.307^{a,b}$ (0.031)
P. freudenreichii ATCC 13673**	9.63 x 10^{8e} (1.11)	$0.06^{c,d}$ (0.01)	$4.95^{a,b}$ (0.01)	0.835^{a} (0.126)	$0.396^{a,e}$ (0.060)
P. jensenii LET128	8.85 x 10^{8e} (0.65)	0.04^{d} (0.00)	5.05^{a} (0.04)	$0.624^{a,b}$ (0.051)	0.915^{d} (0.075)

* Propionibacterium. [a, b, c, d, e] Statistically significant differences between groups. **References strains.

obtained by mechanical disruption. Active cultures of PAB (5 mL) were harvested by centrifugation (10,000 × g, 15 min, 4 °C), washed and suspended to the original volume in 0.9% NaCl and incubated at 37 °C for 2 h to deplete intracellular reserves. These suspensions were inoculated at a 2% (v/v) rate in a basic MRS fermentation broth without glucose and meat extract (Pronadisa), enriched with 0.2% Tween 80, 0.8% casein acid hydrolysate and 0.05% L-cysteine, and supplemented with 0.5% lactose as carbon and energy source to induce β-gal synthesis. After 24 h of incubation at 37 °C, the bacterial cells were harvested by centrifugation (10,000 × g, 15 min, 4 °C), washed and resuspended in 500 μL of 50 mM sodium phosphate buffer, pH 6.5, and mixed (1:1, w/v) with sterile glass beads (diameter, 150–212 mm; Sigma-Aldrich, St. Louis, MO, USA). The cells were disrupted in FastPrep equipment (Bio101; Savant Instruments, Holbrook, NY, USA) through physical beating, and then, the insoluble fraction and glass beads were removed by centrifugation (12,000 × g, 10 min, 4 °C). The supernatant fraction (CFE) containing β-gal for all of the assays was stored at −80 °C until ready for use. The protein contents of CFE were determined by using a Bradford Bio-Rad protein kit (Bio-Rad, Munich, Germany) and bovine serum albumin for preparation of a standard curve. The β-gal activity was determined by measuring the hydrolysis rate of the synthetic substrate *o*-nitrophenyl-β-D-galactopyranoside (ONPG) (Sigma-Aldrich) with a colorimetric method (Hartley and Vedamuthu, 1975). The enzymatic reaction was conducted in 96-well microplates by mixing 10 μL of CFE with 200 μL of 1 mg mL^{-1} ONPG (3.2 mM final concentration) and dissolved in 50 mM sodium phosphate buffer at pH 6.5 containing 1 mM $MgCl_2$. The mixtures were incubated for 10 min at 40 °C, and the amount of released *o*-nitrophenol (ONP) was monitored continuously by measuring absorbance at 420 nm in an automated microplate spectrophotometer (Varioskan Flash, Thermo Fisher Scientific, Waltham, MA, USA). Enzymatic activities were expressed as U mL^{-1}, where one unit of β-gal was defined as the amount of enzyme that released 1 μmol of ONP per minute under defined assay conditions. The specific activity was calculated based on the mg of protein present in the CFE.

2.4. Enzymatic synthesis of oligosaccharides from lactose (GOS) and lactulose (OsLu)

The capacity of PAB β-gal to synthesize oligosaccharides was assessed by incubating CFE of *P. acidipropionici* LET 120 and Lactozym with solutions of lactose and lactulose. The reactions were performed in a final volume of 1.0 mL in microtubes by mixing the enzyme and substrates at the following conditions: pH 6.5, 45 °C, 300 g L^{-1} final concentration of lactose or lactulose, and 1.3 U mL^{-1} of enzyme. Aliquots (200 μL) were withdrawn at specific time intervals (1, 3, 5, 7 and 24 h) and immediately immersed in boiling water for 2 min to inactivate the enzyme. The samples were stored at −80 °C for subsequent analysis. All of the reactions were conducted in duplicate, and the corresponding analytical measurements were completed twice for each enzymatic synthesis condition.

2.5. Growth of P. acidipropionici LET 120 on lactose, lactulose and their derived oligosaccharides

P. acidipropionici LET 120 cultured 24 h at 37 °C in MRS was harvested by centrifugation (10,000 × g for 10 min at 4 °C), washed twice and resuspended in sterile saline solution (0.85%). These suspensions were inoculated (2%) in basic MRS broth supplemented with the tested carbohydrates (lactose, lactulose, Vivinal-GOS and *A. oryzae*-OsLu) at a final concentration of 1% and then pipetted in triplicate into 300-μL wells of sterile 96-well microplates with a lid (Sarstedt Inc., Newton, USA). The glucose and the medium without any carbon source were used as growth controls. The plates were incubated at 37 °C for 48 h in an automated microplate reader (Varioskan Flash), and the optical densities at 600 nm (OD_{600}) were recorded at 60 min intervals. The maximum growth rates (μ_{max}) of *P. acidipropionici* LET 120 were calculated by fitting the curves to a sigmoid model using Microcal Origin 6.0 (OriginLab Corporation, Northampton, MA, USA). The samples were taken at 0, 24 and 48 h for pH, bacterial enumeration and short chain fatty acids (SCFA) determinations. Colony-forming units (CFU) were counted after plating tenfold diluted samples on MRS agar incubated for 120 h at 37 °C in anaerobiosis (Gas-Pack, Anaerogen; Oxoid Ltd., Hampshire, England).

2.6. Determination of the prebiotic capacity of GOS and OsLu synthesized by PAB on probiotic cultures

Recognized strains, such as *B. animalis* subsp. *lactis* BB-12 and *L. casei* CRL 431, were selected for the evaluation of the prebiotic capacity of GOS and OsLu synthesized by *P. acidipropionici* LET 120 from lactose and lactulose, since these microorganisms are representative of the intestinal microbiota and proven probiotics (Aragon et al., 2014; Jungersen et al., 2014). PAB-oligosaccharides were partially purified by selective adsorption onto activated charcoal followed by yeast fermentation to decrease the monosaccharide and lactose contents (Hernandez et al., 2009). In brief, 1 mL of PAB-oligosaccharides formed after 24 h reactions (carried out as described in Section 2.4), and 3 g of activated charcoal (100–400 mesh, Sigma–Aldrich), was mixed in 100 mL of 10% ethanol and stirred for 30 min. The mixtures were filtered through Whatman No. 1 paper under a vacuum and washed with 25 mL of the same ethanolic solution. Desorption of the oligosaccharides from the activated charcoal was conducted by stirring with 100 mL of 50% ethanol for 30 min and further filtration (as before). The filtrates were evaporated under a vacuum at 40 °C, dissolved in 1 mL of deionized water and mixed with 1 mL of *Saccharomyces cerevisiae* (Levex®, Lesaffre, Argentina) in PBS (≅ 10^9 CFU mL^{-1}). Fermentation was conducted with agitation at 37 °C for 24 h, and the mixtures were finally filtered through 0.22 μm filters (Millipore Corp.) to remove the yeasts before using them as carbon sources in the assays. Chromatography analysis (which was performed as described in Section 2.7.1) showed that the procedure removed 100% of the monosaccharides and ≥95% of the disaccharides. Probiotic strains that were grown for 24 h in MRS with 0.05% L-cysteine hydrochloride (MRS-cys) were washed with 0.9% NaCl and inoculated (initial OD_{600} ≅ 0.05) in basic MRS-cys broth supplemented with Vivinal-GOS, *A. oryzae*-OsLu, LET 120-GOS or LET 120-OsLu at 1% and as the sole carbohydrate source. *P. acidipropionici* LET 120 and *E. coli* C3 were tested for comparison. Probiotic cultures and PAB were incubated at 37 °C for 24 h under anaerobiosis, while *E. coli* was incubated in aerobiosis. The cell growth was determined based on the increase in bacterial biomass measured by turbidimetry at 600 nm and the change in pH of the culture media. The prebiotic activity score (PAS) was calculated according to the following equation of Huebner et al. (2007).

$$PAS = \frac{PP_{24} - PP_0}{PG_{24} - PG_0} - \frac{EP_{24} - EP_0}{EG_{24} - EG_0}$$

where PP_{24} and PP_0 are the probiotic biomass (OD_{600}) after 24 h and 0 h of fermentation of selected prebiotics, respectively. PG_{24} and PG_0 are the probiotic OD_{600} on glucose after 24 h and 0 h fermentation, respectively. EP_{24}/EP_0 and EG24/EG0 are the *E. coli* biomass on prebiotics and glucose after 24 h and 0 h fermentation, respectively. Each assay was replicated twice.

2.7. Analysis of GOS and OsLu

2.7.1. HPLC with refractive index detection (HPLC-RID)

The samples from the enzymatic synthesis were diluted with acetonitrile/water (50:50 v/v) at a total carbohydrate concentration of ~10 mg mL^{-1}, filtered onto 0.2 mm pore size PVDF membranes (Millipore, Massachusetts, USA) and analyzed in an HPLC system (equipped with Smartlinepump 100, refractive index detector K-2301 and smart line autosampler AS 3800, Knauer, Germany). The separation

of carbohydrates was conducted on a Kromasil® column (100-NH_2; 250 mm × 4.6 mm, 5 μm particle size) (Akzo Nobel, Brewster, NY, USA) using acetonitrile/water (70:30 v/v) as the mobile phase and eluted in the isocratic mode at a flow rate of 1 mL min^{-1} for 30 min. The injection volume was 20 μL (~800 μg of total carbohydrates). Data acquisition and processing were performed using EuroChrom for Windows Basic Edition v.3.05 software. The carbohydrates in the reaction mixtures were initially identified by comparing their retention times (t_R) with those of standard sugars. Quantitative analysis was performed by the external standard method using calibration curves of each pure standard (including galactose and fructose for monosaccharides, lactose and lactulose for disaccharides, raffinose for trisaccharides and stachyose for tetrasaccharides quantification) in the range 0.05–5 mg mL^{-1}. All of the analyses were performed in duplicate, and relative standard deviation (RSD) values below 10% were obtained in all cases. The amounts of different carbohydrates present in the reaction mixtures were expressed as the percentage of the total carbohydrate content (w/w).

2.7.2. Gas chromatography with flame ionization detector GC-FID

The composition of the GOS and OsLu mixtures synthesized by *P. acidipropionici* LET 120 and Lactozym from lactose and lactulose was determined by GC-FID. The carbohydrates were analyzed as trimethyl silylated oximes (TMSO). First, 15 μL of the reaction mixture (4.5 mg of sugars) was added to 0.4 mL of the internal standard (I.S.) solution (0.5 mg mL^{-1} phenyl-β-glucoside). Afterwards, the mixture was dried at 38–40 °C in a rotary evaporator (Büchi Labortechnik AG, Flawil, Switzerland). The sugar oximes were formed by adding 250 μL of hydroxylamine chloride (2.5%) in pyridine and heating the mixture at 70 °C for 30 min and then silylating with hexamethyldisilazane (250 μL) and TFA (25 μL) and kept at 50 °C for 30 min. The reaction mixtures were centrifuged at 10,000 rpm for 2 min at room temperature. The supernatants were injected or stored at 4 °C prior to analysis.

Gas chromatography (GC) analysis was performed on an Agilent Technologies 7890A gas chromatograph (Wilmington, DE, USA) equipped with a commercial fused silica capillary column DB-5HT, bonded, crosslinked phase (30 m × 0.25 mm i.d. and 0.25 μm film thickness) (J&W Scientific, Folson, California, USA). The oven temperature was initially 180 °C increased at a rate of 3 °C min^{-1} to 350 °C and then held for 25 min. The injector and detector temperatures were set at 280 and 355 °C, respectively. The injections were conducted in split mode (1:30) using nitrogen at 1 mL min^{-1} as the carrier gas. Data acquisition and integration were performed using Agilent ChemStation Rev. B.03.01 software. The response factors were calculated relative to the I.S. using solutions containing galactose, glucose, lactose, raffinose and stachyose and were prepared over the expected concentration range in the samples. The identities of the carbohydrates were confirmed by comparison with relative retention times of standard samples. Bimuno® (Clasado Ltd., Reading United Kingdom), with 65% GOS formed mainly by β-(1 → 3), as well as β-(1 → 4) and β-(1 → 6) (Gibson et al., 2011); and 4′-galactosyl-lactose previously synthesized and identified in our laboratory (Cardelle-Cobas, 2009), were used as standards to tentatively identify GOS synthetized by PAB β-gal. All of the analyses were conducted in duplicate, and the data were expressed as (% total carbohydrates w/w) mean ± standard deviation (SD).

2.8. Analysis of short chain fatty acids (SCFA)

Determination of SCFA originated during the growth of *P. acidipropionici* LET 120 on carbohydrates (lactose, lactulose, Vivinal-GOS and *A. oryzae*-OsLu) was conducted as previously described (Zárate et al., 2017). Before analysis, the culture supernatants were deproteinized with 0.01 mol L^{-1} H_2SO_4 (15 min, 4 °C), centrifuged for 10 min (10,000 × g and 4 °C) and filtered using a 0.45 μm syringe filter. The separation of SCFA were conducted in an HPLC system equipped with Smartlinepump 100, refractive index detector K-2301 and smart line autosampler AS 3800 (Knauer, Germany). The elution of SCFA was performed using a Rezex ROA organic acids column (300 × 7.8 mm and 8 μm particle size) (Phenomenex Torrance, USA) thermostatized at 55 °C. The separation was in the isocratic mode, and the mobile phase was 0.01 M H_2SO_4 at a flow rate of 0.6 mL min^{-1}. Quantification of the organic acids was conducted by the external standard method (Snyder et al., 1997) using standard solutions of acetic acid and propionic acid prepared in concentrations ranging from 0.05 to 5.0 mg mL^{-1}. The linear regression curves for each standard were higher than 0.99. The product concentrations were reported as mg mL^{-1} of grown culture medium.

2.9. Statistical analysis

ANOVA tests and Tukey's test for $p < 0.05$ were applied to all of the data generated including ONPG hydrolysis by β-gal enzymes contained in the CFE of dairy PAB, carbohydrates released with β-gal of *P. acidipropionici* LET 120 and Lactozym, the growth parameters of *P. acidipropionici* LET 120, and probiotic strains at the expense of different carbohydrates such as the carbon source. In addition, a Pearson correlation coefficient was calculated for all of the variables. The differences between the cultures and the substrates employed were studied. All of the statistical analyses were performed on R (R Core Team, 2017).

3. Results and discussion

3.1. Assessment of the β-gal activity of dairy PAB

At present, non-digestible oligosaccharides (NDO), such as GOS, have gained great interest for health and industrial applications, since they are well-recognized prebiotics and food additives (Vera et al., 2016; Zárate et al., 2013). In this regard, it is proposed that β-gal from probiotic strains would be more appropriate for the synthesis of selective prebiotic oligosaccharides (Osman et al., 2010; Rabiu et al., 2001). Then, as a first step in the evaluation and selection of PAB with suitable β-gal for GOS synthesis, we screened the growth and enzyme activity of 17 strains of dairy PAB at the expense of lactose as the sole carbon source. Eleven strains, including the references *P. freudenreichii* ATCC 13673 and *P. acidipropionici* ATCC 22562, expressed β-gal and were able to develop on lactose (*P. acidipropionici,* n = 5; *P. freundereichii,* n = 3; and *P. jenseni,* n = 1 strains). As a general result, *P. acidipropionici* strains grew faster, showed statistically significant differences, and attained higher cell counts and lower pH than the *P. freudenreichii* and *P. jensenii* strains (Table 1).

The β-gal activity of microorganisms quantified on CFE by determining their specific activities varied widely between 0.081 and 2.553 U mg^{-1} for *P. acidipropionici* LET 116 and LET 120, respectively. However, there has been less variation observed in specific activities of *P. freudenreichii strains* (0.214–0.396 for LET 114 and ATCC 13673, respectively) (Table 1). The statistical analysis of the parameters reported in Table 1 shows that maximum cell count, growth rate and the final pH were strongly correlated.

Fig. 1 shows three kinetics of ONPG hydrolysis; most of the strains were distributed among a low and intermediate rate of hydrolysis, whereas only *P. acidipropionici* ATCC 25562 and LET 120 displayed high hydrolysis rates. Statistically significant differences were found between *P. acidipropionici* LET 120 activity and the rest of the strains. Similarly, *P. acidipropionici* ATCC 25562 activity was also significantly different from the rest, whereas no statistically significant differences were observed between strains showing an intermediate rate of hydrolysis (*P. freudenreichii* LET114, LET125 and LET126 and *P. jenseni* LET128) and between strains with a low rate of hydrolysis (*P. acidipropionici* LET113, LET116, LET117 and LET119). These data could indicate that the hydrolysis rate of ONPG does not depend on the

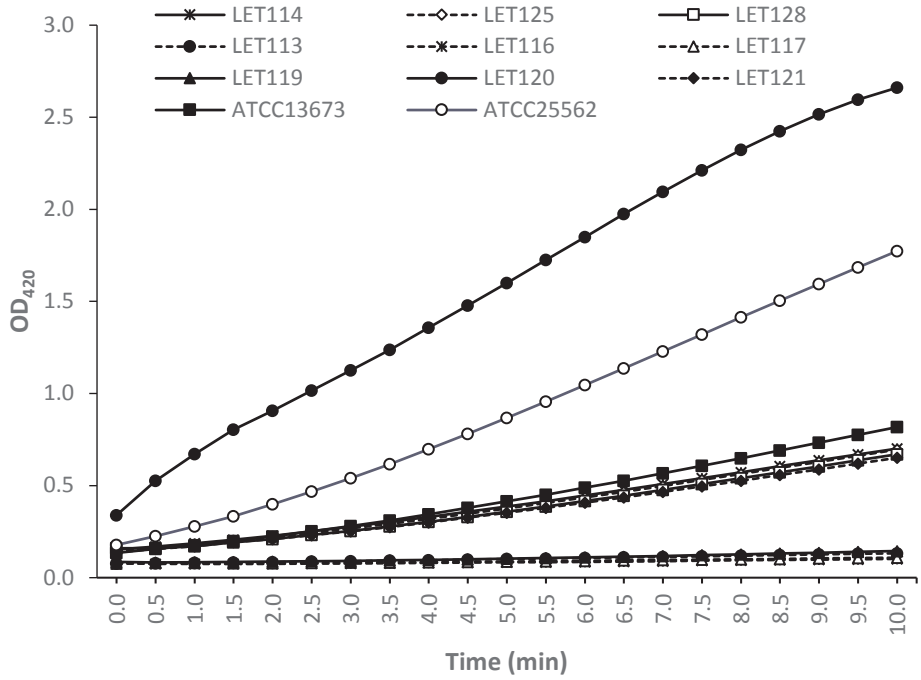


Fig. 1. Monitoring of ONPG hydrolysis (pH 6.5, 40 °C) by the β-galactosidase enzyme contained in the CFE of dairy PAB (codes in Table 1). The data are the means of two determinations of two independent assays (n = 4).

species but on the strain. Since *P. acidipropionici* LET 120 produced a high amount of β-gal with the highest specific activity, it was selected to assess its ability to synthesize GOS and OsLu.

3.2. Synthesis of GOS and OsLu by P. acidipropionici LET 120

Lactose has been mainly used as substrate for the production of GOS (Cardelle-Cobas et al., 2008b, 2009b); however, many studies have proven that other carbohydrates, such as lactulose, are also good substrates for the enzymatic synthesis of prebiotic oligosaccharides using microbial β-gal (Cardelle-Cobas et al., 2008a; Guerrero et al., 2011, 2015; Hernández-Hernández et al., 2011; Martínez Villaluenga et al., 2008b). The CFE of *P. acidipropionici* LET 120 containing the β-gal activity enzyme (2.5 U mL^{-1}) was mixed with lactose or lactulose at a final concentration of 300 g L^{-1} and incubated at 45 °C at a pH 6.5 for 24 h, and samples were taken at intervals for carbohydrate analysis. For a comparison, reactions were also performed with the commercial preparation Lactozym that contains β-gal from *Kluyveromyces lactis*. Figs. 2 and 3 show the time course of β-gal-catalyzed reactions with lactose and lactulose as the substrate, respectively. The oligosaccharides in the reaction mixtures were quantified by HPLC-RID. The reactions with both disaccharides produced, besides hydrolysis into their glucose and galactose monosaccharides (Fig. 2b), the synthesis of GOS by transgalactosylation. As can be observed in Fig. 2a, a similar behavior towards lactose was observed with both enzymes, although the purified enzyme of *K. lactis* was more efficient than the unpurified extract containing β-gal of *P. acidipropionici* LET 120, since it degraded more lactose and released a higher amount of products. The lactose concentration decreased rapidly during the first hours of the reaction, which resulted in GOS (di-, tri- and tetrasaccharides) in concentrations that increased with the progress of the reaction up to 24 h (Fig. 2a). However, the β-gal of *P. acidipropionici* LET 120 produced in the first hour of the reaction almost the maximum levels of GOS, which increased slightly in the following hours. In the optimal conditions using Lactozym, lactose hydrolysis, the release of monosaccharides and GOS formation were significantly higher. Maximum GOS formation by β-gal of *P. acidipropionici* LET 120 was achieved after 24 h of reaction and accounted for 23.6% (w/w) of total sugars, whereas only 5 h of reaction was needed to reach a maximum GOS formation of 25.0% (w/w) of the total sugars with Lactozym. Other studies on the transgalactosylation of the lactose solutions by β-gal from other bacteria, such as *L. reuteri* and *Bifidobacterium* species using different reaction conditions, reported yields of GOS ranging from 26.8 to 47.6% (Rabiu et al., 2001; Splechtna et al., 2006; Hsu et al., 2007; Iqbal et al., 2010).

Lactulose was also hydrolyzed by both β-gal into galactose and fructose (Fig. 3b) and transgalactosylated into OsLu (di- and tri-saccharides). Similar curves of lactulose degradation and production of OsLu were observed for *P. acidipropionici* LET 120 β-gal and Lactozym (Fig. 3a). The maximum production of LET-120 OsLu was achieved after 5 h of reaction and corresponded to 27.0% of the total carbohydrates, whereas a maximum OsLu of 30.5% was obtained with Lactozym at 24 h, and no significant differences were found. However, in OsLu synthesis reactions, lactulose hydrolysis was significantly lower with *P. acidipropionici* LET 120. Although lactose hydrolysis was significantly higher than that of lactulose, a significantly higher amount of oligosaccharides was obtained when lactulose was used as the substrate.

To better understand the oligosaccharides formed in enzymatic hydrolysates of lactose and lactulose using β-gal from *P. acidipropionici* LET 120 and Lactozym, GC-FID analyses of the reaction mixtures at optimal conditions for GOS and OsLu formation were performed. Additionally, a comparison with the GC-FID profile of oligosaccharides found in hydrolysates using β-gal of *K. lactis* was conducted. Thus, different profiles were obtained with both β-gal and a good separation of mono-, di-, tri- and tetrasaccharides (Fig. 4a). The compounds formed during lactose reactions with β-gal from *P. acidipropionici* LET 120 were compared with GOS synthesized by our research group in previous studies (Martínez Villaluenga et al., 2008a; Cardelle-Cobas, 2009) and Bimuno, which is a commercial GOS product. Thus, the disaccharides present in the reaction mixtures were identified as allolactose (β-D-Gal*p*-(1 → 6)-D-Glu; peak 7) and 6-galactobiose (β-D-Gal*p*-(1 → 6)-D-Gal; peak 9). The main trisaccharide (peak 11) was coincident with that previously reported by Martínez-Villaluenga et al. (2008b) as 6′-galactosyl-lactose (*β*-D-Gal*p*-(1 → 6)-lactose). Other trisaccharides were also detected, but only two of them were identified as 3′-galactosyl-lactose (*β*-D-Gal*p*-(1 → 3)-lactose) (peak 12), which is the

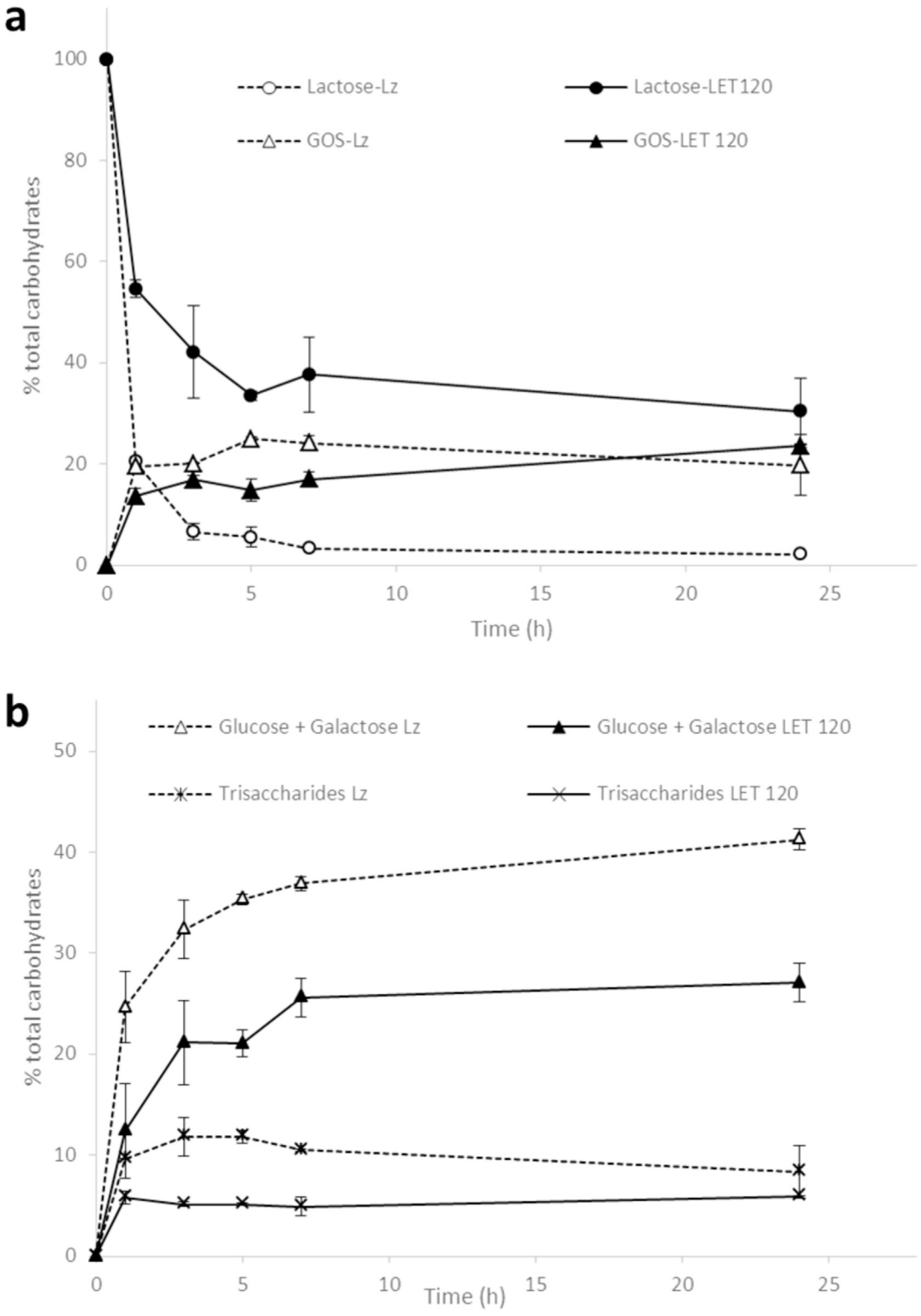


Fig. 2. Time course of the reactions catalyzed with the β-gal of *P. acidipropionici* LET 120 (LET 120) and Lactozym Pure 6500L (Lz) at 45 °C, pH 6.5 and 300 g L^{-1} of lactose as substrate: a) lactose and galactooligosaccharides (GOS) b) mono- and trisaccharides. Carbohydrates in the reaction mixtures were quantified by HPLC-RID.

most abundant trisaccharide found in commercial GOS-Bimuno, and 4′-galactosyl-lactose (peak 10) (Cardelle-Cobas, 2009). Additionally, tetrasaccharides were detected, but none of them could be identified.

Oligosaccharides formed during the lactulose reaction with β-gal of *P. acidipropionici* LET 120 were also studied. As is shown in Fig. 4b, di-, tri- and tetrasaccharides were detected in the GC profile. By comparison with the OsLu synthesized by our research group in previous studies (Martínez Villaluenga et al., 2008b), disaccharides, such as allolactulose (*β*-D-Gal*p*-(1 → 6)-D-Fru) (peaks 8) and 6-galactobiose (peaks 13), could be identified as well as the trisaccharide 6′-galactosyl-lactulose (*β*-D-Gal*p*-(1 → 6)-lactulose). Other trisaccharides as well as tetrasaccharides were also detected, but they could not be identified.

Oligosaccharide quantification in enzymatic mixtures using β-gal from *P. acidipropionici* LET 120 and Lactozym was also conducted with

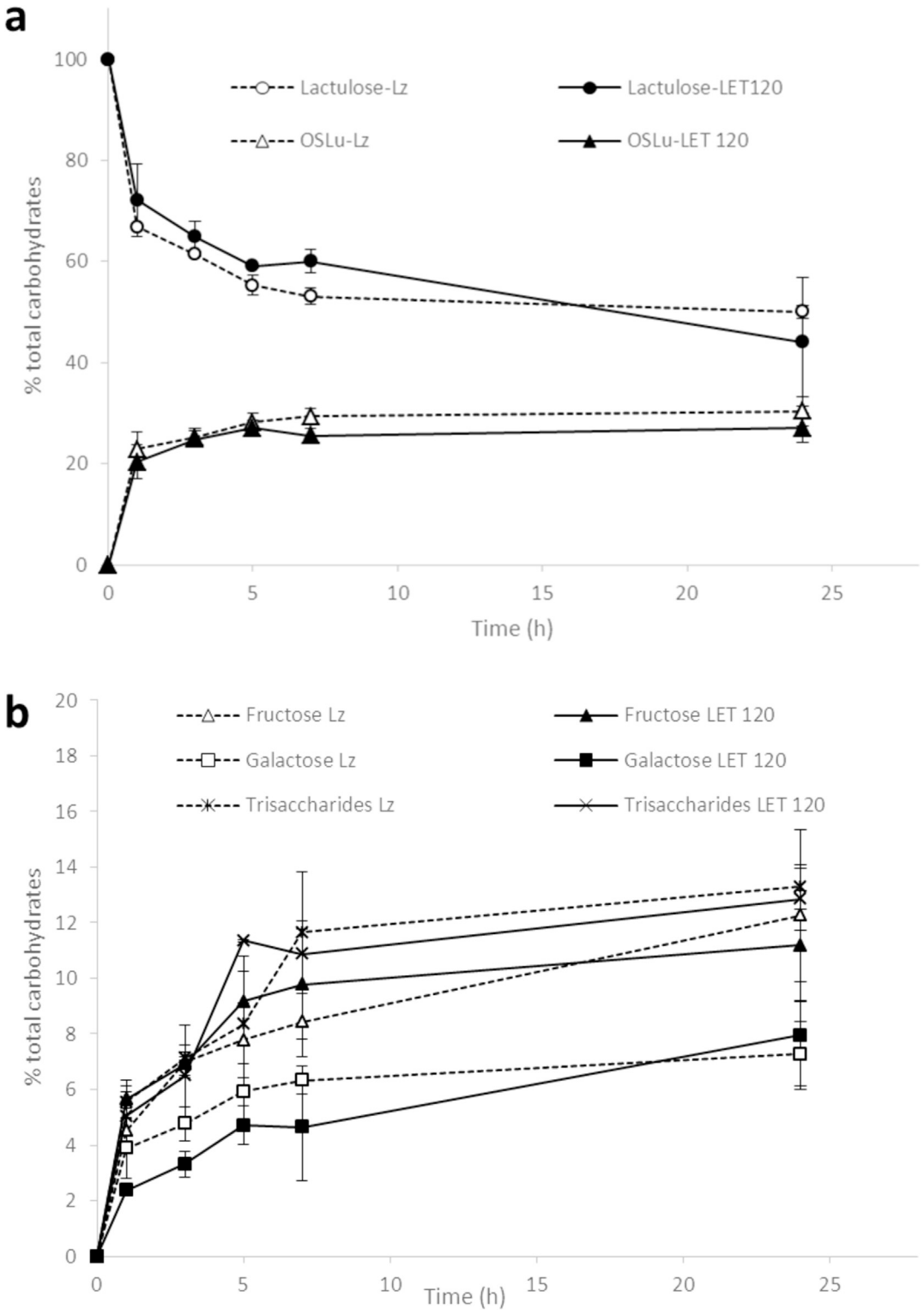
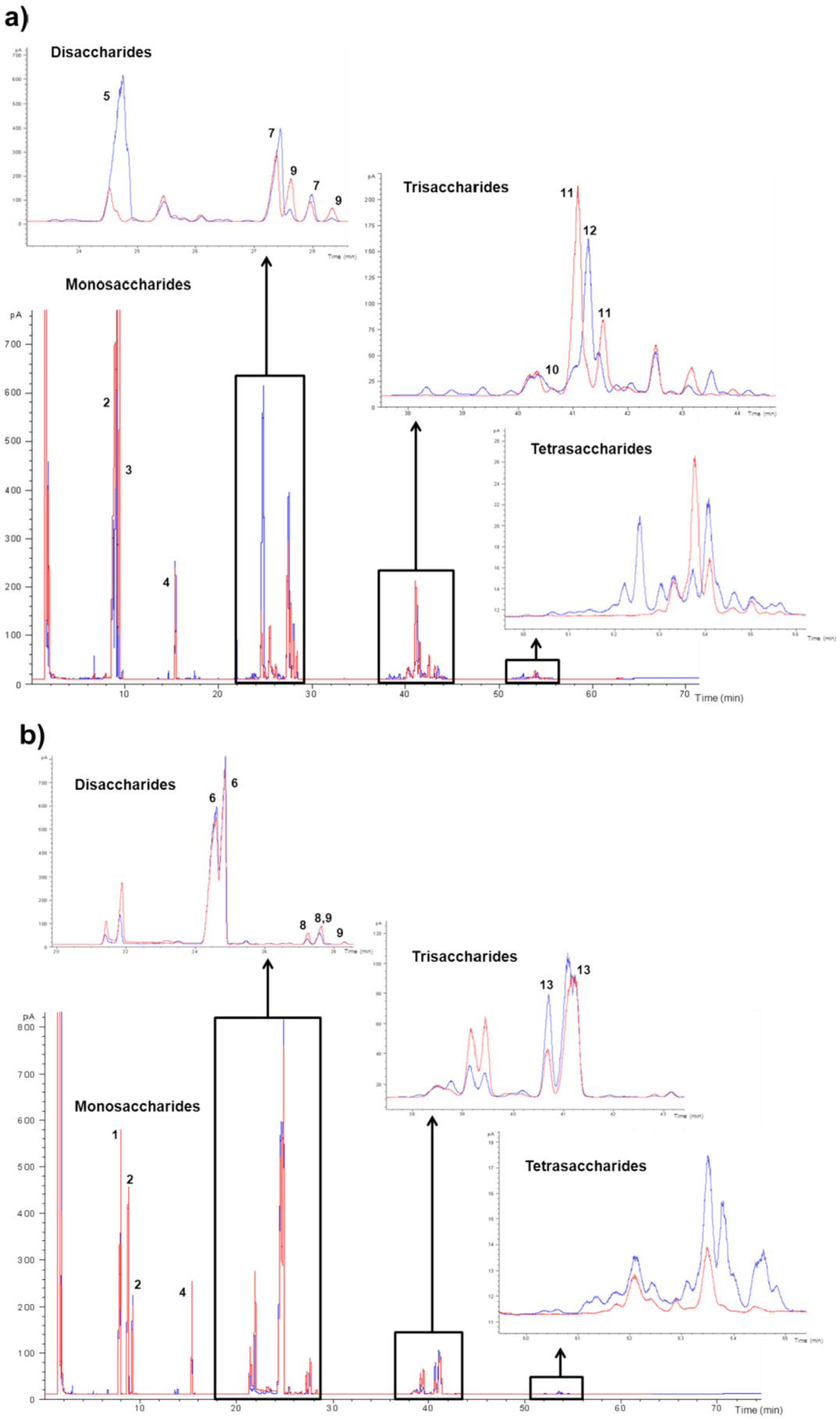


Fig. 3. Time course of the reactions catalyzed with the β-gal of *P. acidipropionici* LET 120 (LET 120) and Lactozym Pure 6500L (Lz) at 45 °C, pH 6.5 and 300 g L^{-1} of lactulose as substrate: a) lactulose and oligosaccharides derived from lactulose (OsLu) b) mono- and trisaccharides. Carbohydrates in the reaction mixtures were quantified by HPLC-RID.

CG-FID. Table 2 shows the maximum value attained for oligosaccharides (di-, tri- and tetrasaccharides) in the two studied hydrolysates containing lactose and lactulose. In hydrolysates with lactose, the main disaccharide formed was allolactose followed by 6-galactobiose. In the case of hydrolysate from β-gal of *P. acidipropionici* LET 120, three trisaccharides were quantified: 6′-galactosyl-lactose, 3′-galactosyl-lactose and 4′-galactosyl-lactose. However, 3′-galactosyl-lactose was not detected in hydrolysates from Lactozym. Unknown di-, tri- and tetrasaccharides were also quantified, and all of them were included in the GOS values (Table 2).

In hydrolysates from lactulose, quantification was also performed by GC-FID (Table 2). The main disaccharide formed was allolactulose



(caption on next page)

Fig. 4. Gas chromatographic (GC-FID) profile of the carbohydrates present in the reaction mixtures of β-galactosidase (1.3 U mL^{-1}) from *P. acidipropionici* LET 120 (blue line) and Lactozym (red line) with 300 g L^{-1} of (a) lactose or (b) lactulose performed at 45 °C, pH 6.5 during 7 h. 1: fructose, 2: galactose, 3: glucose, 4: internal standard, 5: lactose, 6: lactulose, 7: allolactose, 8: allolactulose, 9: 6-galactobiose, 10: 4′-galactosyl lactose, 11: 6′-galactosyl lactose, 12: 3′-galactosyl lactose, 13: 6′-galactosyl lactulose. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

followed by 6-galactobiose. Only a trisaccharide, 6′-galactosyl-lactulose, was quantified in the two hydrolysates. Unknown di-, tri- and tetrasaccharides were also quantified, and all of them were included in the OsLu values (Table 2). It has been previously reported that β-gal from potential probiotic genera, such as *Lactobacillus*, mainly produce β-(1 → 6) linkages in their transgalactosylation mode, whereas β-(1 → 3) linked products are formed to a lesser extent (Splechtna et al., 2006; Iqbal et al., 2010). Depeint et al. (2008) found that GOS mainly formed with β-(1 → 3) linkages using β-gal of *B. bifidum* NCIMB 41171. Accordingly, it has been demonstrated that linkages β-(1 → 6) are preferred and cleaved faster than other linkages by β-gal from bifidobacteria (Cardelle Cobas et al., 2011); thus, GOS produced by *P. acidipropionici* LET 120 could exhibit a good prebiotic potential.

3.3. Growth of P. acidipropionici LET 120 on lactose, lactulose and their derived oligosaccharides

Prebiotic GOS are mainly targeted to stimulate the development of Bifidobacteria in the colon. However, they could also beneficially influence the growth of potential probiotic strains of other genera such as *Lactobacillus* and *Streptococcus* (Cardelle-Cobas et al., 2011). We first assessed the ability of *P. acidipropionici* LET 120 to grow at the expense of recognized prebiotics such as Vivinal-GOS, lactulose and *A. oryzae*-OsLu. Fig. 5 shows growth curves of the *P. acidipropionici* LET 120 strain inoculated on the different assayed carbon sources. Statistically significant differences were found between the maximum absorbance reached in all of the studied growth curves with the exception of lactulose and *A. oryzae*-OsLu, where their maximum absorbance values were not significantly different.

Some relevant parameters related to growth are presented in Table 3. No growth was observed in the control culture without a carbohydrate source. In contrast, *P. acidipropionici* LET 120 was able to develop in the five substrates tested and reached a maximum OD_{600} of 1.1 at 13 h of growth and the highest μ_{max} (0.14 h^{-1}) with *A. oryzae*-OsLu as the carbon source. Similar lag times, growth rates and maximum OD_{600} were observed on *A. oryzae*-OsLu and lactulose. When *P. acidipropionici* LET 120 grew at the expense of Vivinal-GOS, a slightly lower μ_{max} (0.10 h^{-1}) and cell density were achieved (OD_{600} 0.94). Finally, lactose and glucose resulted in lower growth rates and biomasses and reached a maximum OD_{600} of 0.78 and 0.62 for each substrate, respectively, with statistically significant differences with respect to *A. oryzae*-OsLu and lactulose. The lowest μ_{max} but longer growth was obtained at the expense of glucose, although the lag time was similar to the observed with the other carbohydrates.

SCFA production was related to the substrate preference and growth on each carbon source as it was much higher in cultures grown on *A. oryzae*-OsLu, Vivinal-GOS and lactulose than in glucose or lactose. Only fermentation of oligosaccharides maintained the molar ratio of 2:1 between propionic and acetic acids that is typical of this bacterial genus. Regarding β-gal, *P. acidipropionici* LET 120 displayed activity in the absence and presence of the different carbon sources tested, and this suggests the constitutive nature of the enzyme in this species as previously reported (Zárate et al., 2003). However, the production of this enzyme was induced by the presence of different carbohydrates, including glucose, with the highest significant activity found on *P. acidipropionici* LET 120 grown at the expense of *A. oryzae*-OsLu, followed by lactulose and Vivinal-GOS (Table 3) in accordance with the results of growth and organic acids.

Correlation analysis (determined by the Pearson coefficient) between the set of parameters (Table 3) shows that the maximum OD_{600} was strongly correlated with final pH, growth rate, β-gal activity and acetic acid production. Similarly, the maximum cell count, growth rate, enzymatic activity and SCFA production were strongly correlated.

Many GRAS microorganisms that are intended as probiotics are combined in foods with prebiotics. The probiotic potential of dairy PAB has been widely reported (Zárate and Pérez Chaia, 2015). However, to exert health benefits, they must persist in the gastrointestinal tract either by their ability to adhere to the intestinal mucosa or proliferate in the lumen. Thus, in addition to specific surface adhesins, the presence of appropriate enzymes for metabolizing NDO may confer them a competitive advantage over other members of the gut microbiota, and the possibility of increasing their number. The results showed that *A. oryzae*-OsLu and lactulose exerted on *P. acidipropionici* LET 120 had a higher stimulating growth effect than more easily usable sugars such as glucose or lactose and enhanced the production of SCFA (propionic and acetic acids). Previous studies have reported the relationship between the oligosaccharide chemical structure and its potential bioactivities (Cardelle-Cobas et al., 2011; García Cayuela et al., 2014; Rabiu et al., 2001).). It has been reported that some *Bifidobacterium* strains preferably ferment tri- and tetrasaccharide GOS over disaccharides, lactose and more simple carbohydrates, such as glucose or galactose (Amaretti et al., 2007 Gopal et al., 2001), and show a degree of polymerization-specific strain preferences (Barboza et al., 2009). In addition, potential probiotic strains of *Streptococcus, Lactobacillus* and *Bifidobacterium* were able to utilize lactulose and pure trisaccharides derived from lactulose and lactose with a general preference towards β-galactosyl residues β(1 → 6) and β(1 → 1) linked over β(1 → 4) linked. In addition, some strains achieved higher cell densities and rates of growth on 6′-galactosyl-lactulose than on 6′-galactosyl-lactose (Cardelle-Cobas et al., 2011).

Most commercially available GOS are mixtures containing monosaccharides, disaccharides and trisaccharides with different linkages. In the case of Vivinal-GOS, it contains predominantly β(1→4) trisaccharides (Chockchaisawasdee et al., 2005), whereas OsLu synthesized with β-gal of *A. oryzae* contains high percentages of 6′-galactosyl-lactulose (Hernandez-Hernandez et al., 2011). The results suggest that *P. acidipropionici* LET 120 also prefer β(1 → 6) linked residues present in *A. oryzae*-OsLu over the β(1 → 4) GOS present in Vivinal-GOS syrup, but further studies are needed to confirm this hypothesis.

3.4. Determination of the prebiotic capacity of GOS and OsLu synthesized by PAB in probiotic cultures

For consideration as prebiotic candidates, oligosaccharides synthesized by PAB should be able to stimulate the growth of beneficial microorganisms present in the gut such as bifidobacteria and lactobacilli. Then, the utilization of LET 120-GOS and LET 120-OsLu by recognized probiotic strains, such as *B. animalis* subsp. *lactis* BB-12 and *L. casei* CRL 431, was assessed to determine their prebiotic and symbiotic potential. Fermentations were conducted in batch cultures with glucose, Vivinal-GOS, *A. oryzae*-OsLu, LET 120-GOS or LET 120-OsLu as the sole carbohydrate source, and the growth of the strains was evaluated by recording the culture optical density and pH at 24 h. For a given sugar to have prebiotic activity, it should be better metabolized than glucose by a test strain. Both probiotics were able to grow on GOS and Oslu produced by LET 120 as evidenced by the increase in OD_{600} and the decrease in pH. As shown in Table 4, the increase in cell density of both probiotics grown on all of the tested GOS was higher than glucose.

Table 2
Carbohydrates content (% of total carbohydrates in the sample) determined by GC-FID, found in the reaction mixtures with 300 g L^{-1} of lactose or lactulose incubated at 45 °C and pH 6.5 with the β-gal of *P. acidipropionici* LET120 and Lactozym (β-gal from *K. lactis*) at maximum oligosaccharides formation times.

Source of β-galactosidase	Substrate	Time (h)	Carbohydrate concentration (% total carbohydrates w/w)												
			Fr	Ga	Gl	La	Lu	AloLa	AloLu	GaBio	3′GaLa	4′GaLa	6′GaLa	6′GaLu	GOS*	OsLu**
***P. acidipropionici* LET 120**	lactose	24	–	12.2^{a} (0.3)	28.4^{a} (2.2)	28.7^{a} (1.4)	–	7.1^{a} (1.0)	–	4.2^{a} (0.8)	2.1^{a} (0.0)	1.1^{a} (0.1)	5.4^{a} (0.2)	–	26.8^{a} (1.6)	–
	lactulose	24	14.8^{c} (0.8)	10.6^{c} (0.8)	–	–	46.8^{c} (5.0)	–	1.2^{c} (0.5)	0.6^{c} (0.2)	–	–	–	7.2^{c} (0.1)	–	26.1^{c} (1.3)
Lactozym	lactose	5	–	22.6^{b} (5.5)	28.8^{a} (2.7)	11.0^{b} (1.4)	–	8.0^{a} (0.5)	–	4.5^{a} (0.5)	–	0.9^{a} (0.2)	12.9^{b} (1.4)	–	36.9^{b} (3.4)	–
	lactulose	24	14.2^{c} (1.3)	8.8^{c} (0.6)	–	–	44.8^{c} (0.9)	–	2.7^{d} (0.5)	1.4^{a,d} (0.1)	–	–	–	5.5^{c} (1.7)	–	32.2^{d} (1.0)

Fr, fructose; Ga, galactose; Gl, glucose; La, lactose; Lu, lactulose; AloLa, allolactose; AloLu, allolactulose; GaBio, 6-galactobiose; 3′GaLa, 3′galactosyl-lactose; 4′GaLa, 4′galactosyl-lactose; 6′GaLa, 6′galactosyl-lactose; 6′GaLu, 6′galactosyl-lactulose.
a, b, c, d Statistically significant differences between rows for lactose (a,b) and lactulose reactions (c, d).
*GOS: Σ AloLa; GaBio; 3′GaLa; 4′GaLa; 6′ GaLa; ukwnown di-, tri- and tetrasaccharides.
**OsLu: Σ AloLu; GaBio; 6′GaLu; ukwnown di-, tri- and tetrasaccharides.

Fig. 5. Growth curves of *P. acidipropionici* LET120 at expense of different carbohydrates lactose, lactulose, Vivinal¯GOS (GOS) and *A. oryzae*-OsLu (OsLu) at 0.5% final concentration. WO/CH: without carbohydrates. Each curve is the average of three replicates from two independent assays.

However, the prebiotic effect of PAB-oligosaccharides was lower than that of Vivinal-GOS and *A. oryzae*-OsLu. No statistically significant differences were observed between the two probiotic strains in most cases. In spite of the higher growth, a lower decrease in pH was observed at the expense of all prebiotics. Otherwise, as expected and even with similar growth, the final pH values obtained with *B. animalis* subsp. *lactis* BB-12 were significantly higher than those obtained with *L. casei* CRL 431. Differences according to the substrate were observed, and in general, the final pH values were higher for these probiotics grown in LET 120-GOS and LET 120-OsLu.

Huebner et al. (2007) established a prebiotic quantitative score (PAS) to describe the extent to which prebiotics support the selective growth of lactobacilli and bifidobacteria. This prebiotic activity assay is based on the change in cell biomass after 24 h of growth of the probiotic strain on 1% prebiotic or 1% glucose relative to the biomass change of a commensal enteric strain grown under the same conditions. Fig. 6 shows the PAS of the tested oligosaccharides on *B. animalis* subsp. *lactis* BB-12, *L. casei* CRL 431 and *P. acidipropionici* LET 120. The growth of enteric *E. coli* C3 on the prebiotics was lower than growth on glucose; thus, all of the GOS demonstrated positive PAS for the probiotics and PAB. Oligosaccharides synthesized by PAB showed lower PAS than commercial Vivinal-GOS and *A. oryzae*-OsLu for the three strains tested; however, no significant differences were observed between the PAS values obtained with the four substrates. The observed trend could be due to their lower content of monosaccharides and lactose after purification.

With respect to the different strains growing on Vivinal-GOS, *A. oryzae*-OsLu, LET 120-GOS and LET 120-OsLu, in general, PAS was significantly lower for *L casei* CRL 431, whereas no statistically significant differences were observed between *B. animalis* subsp. *lactis* BB-12 and *P. acidipropionici* LET 120. Additionally, the highest PAS of oligosaccharides from PAB was observed with the producer LET 120 (1.05 and 1.20 for LET 120-GOS and LET 120-OsLu, respectively). However, no significant differences were observed between PAS values obtained with the four substrates. Other studies previously determined the prebiotic effect of lactose and lactulose-derived oligosaccharides on pure cultures of potential probiotic strains by assessing growth, substrate consumption and/or metabolites production (Cardelle-Cobas et al., 2011; Garcia Cayuela et al., 2014; Hernández et al., 2012). However, PAS can be considered as a simple method for evaluating prebiotics utilization by beneficial bacteria and has been used by other

Table 3
Some relevant parameters related to the growth of *P. acidipropionici* LET 120 and SCFA formation at expense of lactose, lactulose, Vivinal-GOS and *A. oryzae*-OsLu. Data are the means of two independent assays (± SD). The r^2 for fit of data to sigmoid curves ranged from 0.974 to 0.995 for all the carbon sources tested in the study.

	Time for maximum growth	Maximum OD_{600}	CFU mL^{-1}	Final pH	μ (h^{-1})	β-gal (U mg^{-1})	Acetic acid (mg mL^{-1})	Propionic acid (mg mL^{-1})
No sugar	–	0.17^{a} (0.02)	1.5×10^{6a} (0.1)	5.85^{a} (0.04)	–	0.43^{a} (0.09)	–	–
Glucose	28 h	0.62^{b} (0.03)	1.1×10^{9a} (0.2)	4.93^{b} (0.04)	0.03^{a} (0.01)	$1.57^{a,b}$ (0.29)	0.19^{a} (0.05)	0.77^{a} (0.06)
Lactose	13 h	0.78^{c} (0.03)	1.0×10^{9a} (0.1)	4.95^{b} (0.07)	$0.07^{a,b}$ (0.00)	$2.26^{b,c}$ (0.39)	0.12^{a} (0.02)	0.39^{b} (0.04)
Lactulose	17 h	1.05^{d} (0.03)	2.1×10^{9b} (0.5)	4.90^{b} (0.03)	0.12^{b} (0.02)	2.78^{c} (0.15)	0.45^{b} (0.09)	0.72^{a} (0.05)
Vivinal-GOS	15 h	0.94^{e} (0.02)	3.35×10^{9b} (0.2)	5.02^{b} (0.01)	$0.10^{a,b}$ (0.01)	$2.67^{b,c}$ (0.46)	$0.54^{b,c}$ (0.04)	1.10^{c} (0.08)
***A. oryzae*-OsLu**	13 h	1.06^{d} (0.01)	3.5×10^{9b} (0.6)	5.00^{b} (0.06)	0.14^{b} (0.03)	2.95^{c} (0.15)	0.67^{c} (0.04)	1.59^{d} (0.07)

[a, b, c, d, e] Statistically significant differences between groups.

Table 4
Increase in the cell density between time 0 and time 24 h and final pH for pure cultures of selected probiotic strains grown on various oligosaccharide substrates.

	Lactobacillus casei CRL 431		*Bifidobacterium animalis* subsp *lactis* BB-12	
	Increase of biomass*	Final pH**	Increase of biomass	Final pH
Glucose	$0.55 \pm 0.12^{B,b,c}$	$4.04 \pm 0.06^{B,b}$	$0.39 \pm 0.18^{C,c}$	$5.02 \pm 0.01^{A,c}$
Vivinal-GOS	$0.76 \pm 0.10^{A,a}$	$3.74 \pm 0.08^{B,c}$	$0.66 \pm 0.07^{A,a,b}$	$5.74 \pm 0.09^{A,a,b}$
***A. oryzae*-OsLu**	$0.65 \pm 0.06^{A,a,b}$	$4.05 \pm 0.10^{B,b}$	$0.64 \pm 0.11^{A,a,b}$	$5.42 \pm 0.08^{A,b,c}$
LET 120-GOS	$0.60 \pm 0.14^{A,a,b,c}$	$5.14 \pm 0.01^{B,a}$	$0.57 \pm 0.16^{A,a,b,c}$	$6.29 \pm 0.54^{A,a}$
LET 120-OsLu	$0.62 \pm 0.12^{A,a,b}$	$5.05 \pm 0.11^{B,a}$	$0.54 \pm 0.15^{B,b,c}$	$6.13 \pm 0.28^{A,a,b}$

*Determination of biomass in the pure cultures was done by turbidimetry and reported as OD_{600}.
**Culture medium pH at the end of fermentation of probiotics with different carbohydrates as the only carbon source. Initial pH was 6.50 ± 0.11.
[A,B] Statistically significant differences between the two probiotic strains.
[a,b,c] Statistically significant differences between substrates.

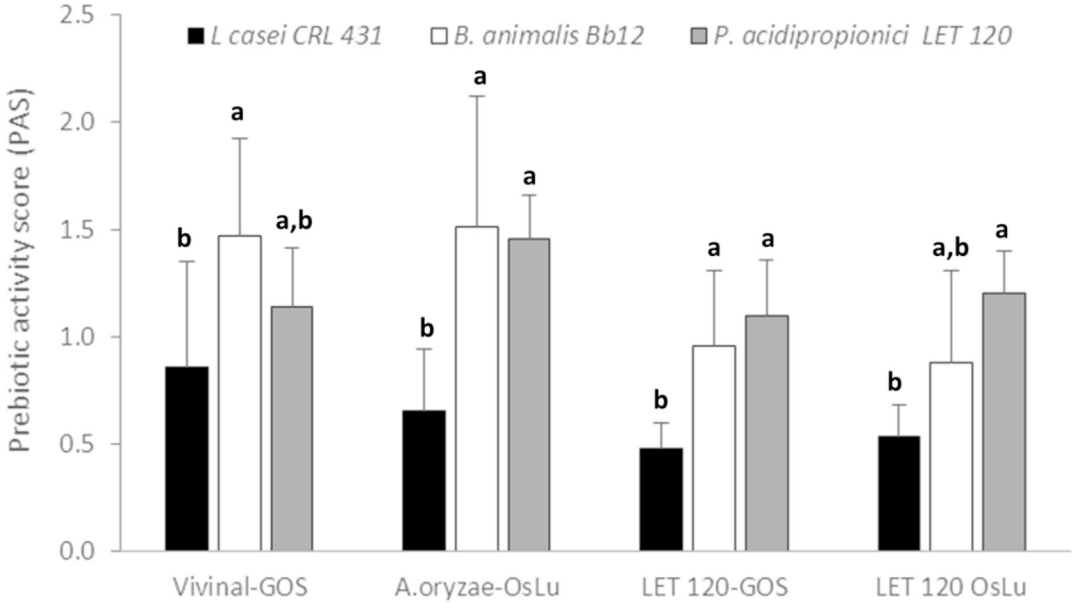


Fig. 6. Prebiotic activity scores (PAS) of oligosaccharides on selected probiotics and *P. acidipropionici* LET 120. ([a,b,c] Statistically significant differences between probiotic strains).

authors as a first approximation to establish the prebiotic condition of oligosaccharides on pure cultures (Guerrero et al., 2015; Huebner et al., 2007; Shi et al., 2018). Additional studies are needed to determine the physiological role of PAB oligosaccharides in the complex gut environment and their potential applications as food ingredients to improve intestinal health.

4. Conclusion

The production of new bioactive oligosaccharides has recently gained much attention for their potential use as functional ingredients. To our knowledge, no studies have addressed the transgalactosidase activity and oligosaccharides synthesis by dairy PAB or their ability to

grow at the expense of NDO. For the first time, the results show that *P. acidipropionici* LET 120 was able to produce GOS β(1 → 6), β(1 → 3) and β(1 → 4) linked trisaccharides and OsLu containing mainly β(1 → 6) linked trisaccharides, which showed prebiotic activity on beneficial *Bifidobacterium* and *Lactobacillus* strains. As additional information, *P. acidipropionici* LET 120 was able to utilize lactulose, GOS and OsLu as carbon sources, which contributes to the knowledge regarding the ability of this potential probiotic bacteria to metabolize these substrates. Further studies to increase insight into the structure of trisaccharides synthesized by *P. acidipropionici* LET 120 and the optimal conditions for their synthesis are in process.

Declaration of interest

None.

Acknowledgments

This work was supported by PICT 2014-3583 from ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica, Argentina), Project I-COOP_2014_COOPB20099 and EMHE-CSIC 2017 MHE-200047 from CSIC (Consejo Superior de Investigaciones Científicas, España); and AGL2017-84614-C2-1-R from MICINN. Carlos Sabater thanks his FPU Predoc contract from Spanish MECD (FPU14/03619).

References

Algieri, F., Rodríguez-Nogales, A., Garrido-Mesa, N., Vezza, T., Garrido-Mesa, J., Utrilla, M.P., Montilla, A., Cardelle-Cobas, A., Olano, A., Corzo, N., Guerra-Hernández, E., Zarzuelo, A., Rodriguez-Cabezas, M.E., Galvez, J., 2014. Intestinal anti-inflammatory effects of oligosaccharides derived from lactulose in the trinitrobenzenesulfonic acid Model of Rat Colitis. J. Agric. Food Chem. 62, 4285–4297.

Amaretti, A., Bernardi, T., Tamburini, E., Zanoni, S., Lomma, M., Matteuzzi, D., Rossi, M., 2007. Kinetics and metabolism of *Bifidobacterium adolescentis* MB 239 growing on glucose, galactose, lactose, and galactooligosaccharides. Appl. Environ. Microbiol. 73, 3637–3644.

Aragón, F., Carino, S., Perdigón, G., de Moreno de LeBlanc, A., 2014. The administration of milk fermented by the probiotic Lactobacillus casei CRL 431 exerts an immunomodulatory effect against a breast tumour in a mouse model. Immunobiology 219 (6), 457–464.

Barboza, M., Sela, D.A., Pirim, C., LoCascio, R.G., Freeman, S.L., German, J.B., Mills, D.A., Lebrilla, C.B., 2009. Glycoprofiling bifidobacterial consumption of galacto-oligosaccharides by mass spectrometry reveals strain-specific, preferential consumption of glycans. Appl. Environ. Microbiol. 23, 7319–7325.

Benavente, R., Pessela, B.C., Curiel, J.A., de las Rivas, B., Munoz, R., Guisan, J.M., Mancheno, J.M., Cardelle-Cobas, A., Ruiz-Matute, A.I., Corzo, N., 2015. Improving properties of a novel β-Galactosidase from *Lactobacillus plantarum* by covalent immobilization. Molecules 20, 7874–7889.

Cardelle-Cobas, A., 2009. Synthesis, Characterization and Prebiotic Properties of Oligosaccharides Derived from Lactulose. Ph D Thesis. Universidad Autónoma de Madrid.

Cardelle-Cobas, A., Corzo, N., Olano, A., Peláez, C., Requena, T., Ávila, M., 2011. Galactooligosaccharides derived from lactose and lactulose: influence of structure on *Lactobacillus, Streptococcus* and *Bifidobacterium* growth. Int. J. Food Microbiol. 149, 81–87.

Cardelle-Cobas, A., Fernandez, M., Salazar, N., Martinez-Villaluenga, C., Villamiel, M., Ruas-Madiedo, P., de los Reyes-Gavilan, C., 2009a. Bifidogenic effect and stimulation of short chain fatty acid production in human faecal slurry cultures by oligosaccharides derived from lactose and lactulose. J. Dairy Res. 76, 317–325.

Cardelle-Cobas, A., Martínez-Villaluenga, C., Sanz, M.L., Montilla, A., 2009b. Gas chromatographic–mass spectrometric analysis of galactosyl derivatives obtained by the action of two different β-galactosidases. Food Chem. 114, 1099–1105.

Cardelle-Cobas, A., Martínez-Villaluenga, C., Villamiel, M., Olano, A., Corzo, N., 2008a. Synthesis of oligosaccharides derived from lactulose and Pectinex Ultra SP-L. J. Agric. Food Chem. 56, 3328–3333.

Cardelle Cobas, A., Olano, A., Irazoqui, G., Giacomini, C., Batista, V.F., Corzo, N., Corzo-Martinez, M., 2016. Synthesis of oligosaccharides derived from lactulose (OsLu) using soluble and immobilized *Aspergillus oryzae* β-Galactosidase. Front. Bioeng. Biotechnol 4, 1–21.

Cardelle-Cobas, A., Villamiel, M., Olano, A., Corzo, N., 2008b. Study of galactooligosaccharides formation from lactose using Pectinex-Ultra SP-L. J. Sci. Food Agric. 88, 954–961.

Chockchaisawasdee, S., Athanasopoulos, V.I., Niranjan, K., Rastall, R.A., 2005. Synthesis of galacto-oligosaccharide from lactose using β-galactosidase from *Kluyveromyces lactis*: studies on batch and continuous UF membrane-fitted bioreactors. Biotechnol. Bioeng. 89, 434–443.

Código Alimentario Argentino, 1969. Alimentos de Regimen o Dietéticos. Cap XVII. www.anmat.gov.ar.

Corzo-Martınez, M., Copoví, P., Olano, A., Moreno, F.J., Montilla, A., 2013. Synthesis of prebiotic carbohydrates derived from cheese whey permeate by a combined process of isomerisation and transgalactosylation. J. Sci. Food Agric. 93, 1591–1597.

Depeint, F., Tzortzis, G., Vulevic, J., I'Anson, K., Gibson, G.R., 2008. Prebiotic evaluation of a novel galactooligosaccharide mixture produced by the enzymatic activity of *Bifidobacterium bifidum* NCIMB 41171, in healthy humans: a randomized, double-blind, crossover, placebo-controlled intervention study. Am. J. Clin. Nutr. 87, 785–791.

EFSA Panel on Biological Hazards (BIOHAZ), 2013. Scientific Opinion on the maintenance of the list of QPS biological agents intentionally added to food and feed (2013 update). EFSA J. 11 (3449). https://doi.org/10.2903/j.efsa.2013.3449.

García-Cayuela, T., Díez-Municio, M., Herrero, M., Martínez-Cuesta, M.C., Peláez, C., Requena, T., Moreno, F.J., 2014. Selective fermentation of potential prebiotic lactose-derived oligosaccharides by probiotic bacteria. Int. Dairy J. 38, 11–15.

Gibson, G.R., Hutkins, R., Sanders, M.E., Prescott, S.L., Reimer, R.A., Salminen, S.J., Scott, K., Stanton, C., Swanson, K.S., Cani, P.D., Verbeke, K., Reid, G., 2017. The International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics. Nat. Rev. Gastroenterol. Hepatol. 14, 491–502.

Gibson, G., Slupinski, J.W., Tzortzis, G., Wynne, A.G., 2011. Galactooligosaccharide composition and the preparation thereof; Patent US 7,883,874 B2. https://patentimages.storage.googleapis.com/05/fd/3d/b85ce99570c097/US7883874.pdf.

Gopal, P.K., Sullivan, P.A., Smart, J.B., 2001. Utilisation of galacto-oligosaccharides as selective substrates for growth by lactic acid bacteria including *Bifidobacterium lactis* DR10 and *Lactobacillus rhamnosus* DR20. Int. Dairy J. 11, 19–25.

Guerrero, C., Vera, C., Plou, F., Illanes, A., 2011. Influence of reaction conditions on the selectivity of the synthesis of lactulose with microbial β-galactosidases. J. Mol. Catal. B Enzym. 72, 206–212.

Guerrero, C., Vera, C., Acevedo, F., Illanes, A., 2015. Simultaneous synthesis of mixtures of lactulose and galacto-oligosaccharides and their selective fermentation. J. Biotechnol. 209, 31–40.

Hartley, J.C., Vedamuthu, E.R., 1975. β-Galactosidase of *Propionibacterium shermanii*. Appl. Microbiol. Biotechnol. 29, 74–80.

Hernández, O., Ruiz-Matute, A.I., Olano, A., Moreno, F.J., Sanz, M.L., 2009. Comparison of fractionation techniques to obtain prebiotic galactooligosaccharides. Int. Dairy J 19, 531–536.

Hernández-Hernández, O., Montañés, F., Clemente, A., Moreno, F.J., Sanz, M.L., 2011. Characterization of galactooligosaccharides derived from lactulose. J. Chromatogr. A. 1218, 7691–7696.

Hernández-Hernández, O., Muthaiyan, A., Moreno, F.J., Montilla, A., Sanz, M.L., Ricke, S.C., 2012. Effect of prebiotic carbohydrates on the growth and tolerance of Lactobacillus. Food Microbiol. 30, 355–361.

Hill, C., Guarner, F., Reid, G., Gibson, G.R., Merenstein, D.J., Pot, B., Morelli, L., Canani, R.B., Flint, H.J., Salminen, S., Calder, P.C., Sanders, M.E., 2014. The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic. Nat. Rev. Gastroenterol. Hepatol. 11, 506–514.

Hsu, C.A., Lee, S.L., Chou, C.C., 2007. Enzymatic production of galactooligosaccharides by β-galactosidase from *Bifidobacterium longum* BCRC 15708. J. Agric. Food Chem. 55, 2225–2230.

Huebner, J., Wehling, R.L., Hutkins, R.W., 2007. Functional activity of commercial prebiotics. Int. Dairy J. 17, 770–775.

Iqbal, S., Nguyen, T.H., Nguyen, T.T., Maischberger, T., Haltrich, D., 2010. β-Galactosidase from *Lactobacillus plantarum* WCFS1: biochemical characterization and formation of prebiotic galacto-oligosaccharides. Carbohydr. Res. 345, 1408–1416.

Jungersen, M., Wind, A., Johansen, E., Christensen, J.E., Stuer-Lauridsen, B., Eskesen, D., 2014. The Science behind the probiotic strain *Bifidobacterium animalis* subsp. *Lactis* BB-12®. Microorganisms 2 (2), 92–110 28.

López-Sanz, S., Montilla, A., Moreno, F.J., Villamiel, M., 2015. Stability of oligosaccharides derived from lactulose during the processing of milk and apple juice. Food Chem. 183, 64–71.

Lorenzo-Pisarello, M.J., Gultemirian, M.L., Nieto-Peñalver, C., Perez Chaia, A., 2010. *Propionibacterium acidipropionici* CRL 1198 influences the production of acids and the growth of bacterial genera stimulated by inulin in a murine model of cecal slurries. Anaerobe 16, 345–354.

Martínez-Villaluenga, C., Cardelle-Cobas, A., Corzo, N., Olano, A., Villamiel, M., 2008a. Optimization of conditions for galactooligosaccharide synthesis during lactose hydrolysis by β-galactosidase from *Kluyveromyces lactis* (Lactozym 3000 L HP G). Food Chem. 107, 258–264.

Martínez-Villaluenga, C., Cardelle-Cobas, A., Olano, A., Corzo, N., Villamiel, M., Jimeno, M.L., 2008b. Enzymatic synthesis and identification of two trisaccharides produced from lactulose by transgalactosylation. J. Agric. Food Chem. 56, 557–563.

Moreno, F.J., Montilla, A., Villamiel, M., Corzo, N., Olano, A., 2014. Analysis, structural characterization, and bioactivity of oligosaccharides derived from lactose. Electrophoresis 35, 1519–1534.

Osman, A., Tzortzis, G., Rastall, R.A., Charalampopoulos, D., 2010. A comprehensive investigation of the synthesis of prebiotic galactooligosaccharides by whole cells of *Bifidobacterium bifidum* NCIMB 41171. J. Biotechnol. 150, 140–148.

Perez Chaia, A., Zárate, G., 2005. Dairy propionibacteria from milk or cheese diets remain viable and enhance propionic acid production in the mouse cecum. Lait 85, 85–98.

R Core Team, 2017. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rabah, H., Rosa do Carmo, F.L., Jan, G., 2017. Dairy propionibacteria: Versatile probiotics. Microorganisms 5, 1–24.

Rabiu, B.A., Jay, A.J., Gibson, G.R., Rastall, R.A., 2001. Synthesis and fermentation properties of novel galacto-oligosaccharides by β-galactosidases from *Bifidobacterium* species. Appl. Environ. Microbiol. 67, 2526–2530.

Roberfroid, M., Gibson, G.R., Hoyles, L., McCartney, A.L., Rastall, R., Rowland, I., Wolvers, D., Watzl, B., Szajewska, H., Stahl, B., Guarner, F., Respondek, F., Whelan, K., Coxam, V., Davicco, M.J., Leotoing, L., Wittrant, Y., Delzenne, N.M., Cani, P.D., Neyrinck, A.M., Meheust, A., 2010. Prebiotic effects: metabolic and health benefits. Br. J. Nutr. 104, S1–S63.

Shi, Y., Liu, J., Yan, Q., You, X., Yang, S., Jiang, Z., 2018. In vitro digestibility and prebiotic potential of curdlan (1→3)-β-D-glucan oligosaccharides in Lactobacillus species. Carbohydr. Polym. 188, 17–26.

Snyder, L.R., Kirkland, J.J., Glajch, J.L., 1997. Practical HPLC Method Development, second ed. John Wiley & Sons, New York, pp. 643–684.

Splechtna, B., Nguyen, T.H., Haltrich, D., 2006. Production of prebiotic galacto- oligosaccharides from lactose using β-galactosidases from *Lactobacillus reuteri*. J. Agric. Food Chem. 54, 4999–5006.

Urrutia, P., Rodriguez-Colinas, B., Fernandez-Arrojo, L., Ballesteros, A.O., Wilson, L., Illanes, A., Plou, F.J., 2013. Detailed analysis of galactooligosaccharides synthesis with β-galactosidase from *Aspergillus oryzae*. J. Agric. Food Chem. 61, 1081–1087.

Vera, C., Córdova, A., Aburto, C., Guerrero, C., Suárez, S., Illanes, A., 2016. Synthesis and purification of galacto-oligosaccharides: state of the art. World J. Microbiol. Biotechnol. 32 (197), 1–20.

Vera, C., Guerrero, C., Conejeros, R., Illanes, A., 2012. Synthesis of galacto-oligosaccharides by β-galactosidase from *Aspergillus oryzae* using partially dissolved and supersaturated solution of lactose. Enzyme Microb. Technol 50, 188–194.

Vonk, R.J., Reckman, G.A.R., Harmsen, H.J.M., Priebe, M.G., 2012. Probiotics and lactose intolerance. In: Rigobelo, E. (Ed.), Probiotics, https://doi.org/10.5772/51424. InTech. https://www.intechopen.com/books/probiotics/probiotics-and-lactose-intolerance.

Yin, H., Bultema, J.B., Dijkhuizen, L., van Leeuwen, S.S., 2017. Reaction kinetics and galactooligosaccharide product profiles of the β-galactosidases from *Bacillus circulans, Kluyveromyces lactis* and *Aspergillus oryzae*. Food Chem. 225, 230–238.

Zárate, G., Perez Chaia, A., 2015. Propionibacteria also have probiotic potential. In: Venema, K., do Carmo, A.P. (Eds.), Probiotics and Prebiotics: Current Research and Future Trends. Caister Academic Press, Poole, UK, pp. 69–91.

Zárate, G., Pérez Chaia, A., 2012. Influence of lactose and lactate on growth and β-galactosidase activity of potential probiotic *Propionibacterium acidipropionici*. Anaerobe 18, 25–30.

Zarate, G., Perez Chaia, A., Oliver, G., 2003. Some characteristics of practical relevance of the β-Galactosidase from potential probiotic strains of *Propionibacterium acidipropionici*. Anaerobe 8, 259–267.

Zárate, G., Sáez, G., Pérez Chaia, A., 2013. Microbial transformation of lactose: potential of β-galactosidases for probiotic and prebiotic purposes. In: Green, D., Lee, E. (Eds.), Lactose: Structure, Food Industry Applications and Role in Disorders. Nova Science Publishing, USA, pp. 1–50.

Zárate, G., Chaia, A.P., Gonzalez, S., Oliver, G., 2000. Viability and β-galactosidase activity of dairy propionibacteria subjected to digestion by artificial gastric and intestinal fluids. J. Food Prot 63, 1214–1221.

Zárate, G., Sáez, G.D., Pérez Chaia, A., 2017. Dairy propionibacteria prevent the proliferative effect of plant lectins on SW480 cells and protect the metabolic activity of the intestinal microbiota *in vitro*. Anaerobe 44, 58–65.